<table>
<thead>
<tr>
<th>Title</th>
<th>Investigating the use of the Relational Evaluation Procedure (REP) in children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Hayes, Jennifer</td>
</tr>
<tr>
<td>Publication Date</td>
<td>2017-03-27</td>
</tr>
<tr>
<td>Item record</td>
<td><a href="http://hdl.handle.net/10379/6411">http://hdl.handle.net/10379/6411</a></td>
</tr>
</tbody>
</table>

Some rights reserved. For more information, please see the item record link above.
Investigating the use of the Relational Evaluation Procedure (REP) in Children

Jennifer Hayes

BA (Hons) Psychology

MSc Applied Behaviour Analysis

BCBA

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

September, 2016
# Table of Contents

Declaration and Statements ........................................................................................... v

Abstract........................................................................................................................... vi

Acknowledgements ......................................................................................................... viii

Publications and Conference Presentations ................................................................. x

List of Figures and Tables ............................................................................................... xii

Abbreviations .................................................................................................................. xv

Chapter 1 Introduction .................................................................................................... 1

Chapter 2 Behaviour Analytic Approaches to Language ................................................ 9
  Language within Behaviour Analysis ........................................................................... 10
  Verbal Behaviour .......................................................................................................... 11
  Linguistic Generativity and Derived Stimulus Relations ......................................... 17
  Derived Relational Responding (DRR) ...................................................................... 21
  Relational Frame Theory (RFT) .................................................................................. 24
  Summary ..................................................................................................................... 30

Chapter 3 Establishing Derived Relational Responding (DRR) Abilities in Children ................................................................................................................................. 32
  Training Frames of Coordination .............................................................................. 34
  Training Frames of Distinction ................................................................................... 36
  Training Frames of Comparison ................................................................................. 37
  Training Frames of Opposition ................................................................................... 39
  Training Deictic Frames .............................................................................................. 41
  Summary ................................................................................................................... 44

Chapter 4 The Relational Evaluation Procedure (REP) .................................................. 45
  Development of the REP ............................................................................................. 53
  The REP ...................................................................................................................... 55
  REP Research ............................................................................................................. 58
  The REP and the SMART ......................................................................................... 61
  The REP and the IRAP ............................................................................................... 64
  The REP and the RCP ............................................................................................... 67
  REP Research with Children ..................................................................................... 69
  The NSD-REP ............................................................................................................ 71
# Table of Contents

Summary ................................................................................................................. 74

## Chapter 5 Assessment and training of Yes/No Question (Y/N-Q) responding in Children

The Importance of Y/N-Q Responding ................................................................. 76
Y/N-Q in the Cognitive Developmental Literature ................................................. 78
Y/N-Qs in Children with Developmental Delay ...................................................... 80
Theories of Y/N-Q Responding ............................................................................. 82
Summary .................................................................................................................. 87

## Chapter 6 Assessing and Training Relational Evaluation in Young Children using the NSD-REP

Rationale for Studies 1-5......................................................................................... 90
Study 1: Method ..................................................................................................... 92
Results and Discussion ......................................................................................... 98
Study 2 .................................................................................................................. 104
Method .................................................................................................................. 104
Results and Discussion ......................................................................................... 107
Study 3 .................................................................................................................. 109
Method .................................................................................................................. 110
Results and Discussion ......................................................................................... 113
Study 4 .................................................................................................................. 116
Method .................................................................................................................. 116
Results and Discussion ......................................................................................... 119
Study 5 .................................................................................................................. 123
Method .................................................................................................................. 124
Results and Discussion ......................................................................................... 125
General Discussion ............................................................................................... 129
Summary .................................................................................................................. 133

## Chapter 7 Assessing and Training Relational Evaluation in Children with Autism Spectrum Disorder using the NSD-REP

Rationale for Study 6 ............................................................................................. 136
Method .................................................................................................................. 139
Results ................................................................................................................... 144
Discussion ............................................................................................................. 149
Summary .................................................................................................................. 152
# Table of Contents

**Chapter 8 Assessing and Training Relational Evaluation in Mid-Childhood using the NSD-REP**

- Rationale for Studies 7-9 ................................................................. 155
- Study 7: Method .............................................................................. 162
- Results and Discussion ................................................................. 167
- Study 8 .......................................................................................... 174
- Method .......................................................................................... 174
- Results and Discussion ................................................................. 178
- Study 9 .......................................................................................... 181
- Method .......................................................................................... 181
- Results and Discussion ................................................................. 188
- General Discussion ....................................................................... 191
- Summary ....................................................................................... 196

**Chapter 9 Comparing the effects of Derived Relational Training and Computer Coding on Intellectual Potential in School Age Children**

- Rationale for Study 10 ................................................................. 198
- Method .......................................................................................... 200
- Results .......................................................................................... 209
- Discussion ..................................................................................... 216
- Summary ....................................................................................... 221

**Chapter 10 General Discussion**

- Overview of Research Area and Aims of the Thesis ..................... 223
- Summary of Thesis Findings ......................................................... 225
- Theoretical Issues, Implications and Directions for Future Research .... 232
- Summary and Conclusion ............................................................... 250

**References** ................................................................................. 253

**Appendix A: The NSD-REP** ........................................................... 292

**Appendix B: Yes/No Pre-Screening Assessment** ................................. 296

**Appendix B: RAI/SMART Test Instructions** .................................... 297
Declarations Regarding the Work

I, the Candidate, certify that the Thesis is all my own work and that I have not obtained a degree in this University or elsewhere on the basis of any of this work.

This thesis is the result of my own investigations, except where otherwise stated. The following chapter includes work conducted in collaboration with another student:

- Chapter 7, Study 6, was conducted in collaboration with Orla Corbett, an ABA MSc Student.

Signed:........................................... Date:........................................
Abstract
The core aim of this thesis was to explore the use of the Relational Evaluation Procedure (REP) in children, both typically developing, as well as developmentally delayed. The REP (e.g., Stewart, Barnes-Holmes & Roche, 2004) is a methodology based on Relational Frame Theory (RFT; Hayes, Barnes-Holmes & Roche, 2001), a contextual behaviour analytic approach to language and cognition that conceptualises these phenomena in terms of derived relational responding (DRR). In order to analyse and experimentally manipulate DRR efficiently, the REP was developed. This protocol induces participants to evaluate the stimulus relation(s) being presented on a given trial using confirmation or disconfirmation (i.e., “Yes” and “No”, or analogous terms).

Recently, the REP has been used to rapidly train up relational responding using stimuli possessing already established cue functions (e.g., the English words “Yes” and “No”). It has evolved into several forms of protocol (e.g., SMART, IRAP and RCP), useful for researching various dimensions of relational responding. Most work using the REP has been conducted using adults as participants; however given the advantages of this methodology and its potential as an educational methodology, it seems useful to examine its use in children. The purpose of the current work was to investigate the use of the REP in a range of ages of children, both typically developing and developmentally delayed, over the course of several experiments.

Study 1 (n = 26) explored the use of an REP protocol, namely the NSD-REP (Non-arbitrary Same and Different Relational Evaluation Procedure), focused on affirmation and disconfirmation of non-arbitrary same/different relations as an assessment protocol in young typically developing children (i.e., 2.5-4 years). This study involved examining quantitative correlations between the REP protocol, the Preschool Language Scale-4th Edition (PLS4; Zimmerman, Steiner & Pond, 2002) and the Stanford Binet Intelligence Scales-5th Edition (SB5; Roid, 2003). The results showed significant positive correlations between the REP, the PLS4 and the SB5. The second (n = 4), third (n = 8) and fourth (n = 3) studies, involving young typically developing children, explored the use of the NSD-REP as a training protocol. Using a single subject design, REP training was shown to be effective in establishing different levels of affirmation/disconfirmation of non-arbitrary same/different relations in these young participants. Study 5 involved an analysis of the data from Studies 2-4 as a means of providing a preliminary test of the effects of NSD-REP training on the intellectual potential of the participants involved in those studies. The results indicated that participants showed improvement on measures of cognitive and linguistic ability.

Study 6 continued the exploration of the NSD-REP with children on the autism spectrum, aged 5-12 years. Initially 9 children were assessed on the PLS4. Subsequently, three participants were trained on Level 2 of the NSD-REP, before being re-assessed on the
PLS4. The findings illustrated a significant correlation between the NSD-REP and the PLS4. Additionally, all children met criterion within a number of training sessions and this was accompanied by improvements in PLS4 age-equivalent scores in the case of all three participants.

The results of Studies 1-6 provided detailed insights into the NSD-REP and the protocol evolved based on findings from this regime of testing. Using the more refined version of this methodology, Studies 7-9 were conducted with a slightly older pool of typically developing children. Study 7 (n = 23) provided further substantiation of the NSD-REP as an assessment tool, this time with neurotypical children aged 6-7 years. This study explored correlations between the NSD-REP, the SB5, the Relational Abilities Index (RAI) and standardised school assessments (i.e., Math and Reading). The results showed significant correlations between the NSD-REP, SB5 and the math assessment. Studies 8 (n = 3) and 9 (n = 3) advanced this work further by using the NSD-REP to teach children from this older age group and a younger group (i.e., 4 years) to respond to negation controlled questions in the context of non-arbitrary same and different relations.

Using a group design, Study 10 (n = 28) further advanced the use of the REP by using an already established REP computer protocol (i.e., SMART- Strengthening Mental Abilities Using Relational Training) and comparing it with a computer coding programme (i.e., Scratch) on IQ and academically relevant tasks. Significant improvements were seen in measures of overall intellectual performance and in academic attainment, in the SMART group alone, and there was significantly more improvement on these measures in the SMART group than in the Scratch group. These results extend previous work using this REP format to train young children in relational framing.

In summary, the work presented in this thesis is promising for the future development of the REP. It provides information concerning the age at which children could begin to be taught to use the REP and information concerning the development of key skills useful for the REP, including a repertoire of yes/ no responding. The research also indicates the potential of the protocol as a comprehensive tool for assessment and training of relations with both typically developing children and children with ASD. This would constitute a key addition to existing tools for behavioural intervention.
Acknowledgements

“It isn't what we say or think that defines us, but what we do.”
— Jane Austen

It is with immense gratitude that I acknowledge the following people who have encouraged, inspired and supported me throughout this academic journey:

Dr. Ian Stewart- Where to start? From the commencement of this research path to the end, you have been an incredible supervisor, patient, kind and always only an email away! You brought me up a very steep learning path with this PhD and I will always be grateful for what you have taught me, both academically and professionally.

Dr. Jennifer Holloway- A mentor. Your love for Applied Behaviour Analysis and your dedication to striving for improvement have helped me be the best practitioner I can be. You have always been a voice of reason (and kindness) and your work ethic is something I hope I can emulate going forward!

Dr. Sarah Cassidy- For your time! Although exceptionally busy, you took the time to answer every query I had relating to the practical elements of research and have always been so supportive of both me and my research. For this, I am very thankful.

My friends – To all the friends I have met along the PhD path and also a special thanks to Aoife, Catherine, Linda and Eva. I can’t tell you how much I appreciate everything you have done for me over the last five years. For making me laugh when I wanted to cry and lifting me up when I fell down, I could not ask for better friends.

Kris – Can you believe it? I have finally finished! Your kindness, patience and humour have gotten me through the best of times and the worst of times. I could not have finished this PhD without your love and support and I look forward to the next chapter of our story.
All of my participants and the schools that facilitated this research- For giving your time. Without you, this thesis could not have been completed. A special thanks also to the schools and day care facilities, including all the staff who helped me along the way, I will always be so appreciative of your help.

My extended family- For listening to me and encouraging me at every opportunity!

My sincerest gratitude also to:

My sister Taylor and brother Scott- Thank you for all the laughter, especially when I needed it most and for always being my involuntary research assistants!

My mother, Sharon and my sister Aundrea- This academic journey would not have even started without you. You're belief in me gave me the courage I needed to pursue this career path and to continue on it! I owe everything I have achieved to you. I will never forget the love and support you have given me over the last 12 years and I will always be indebted to you. This thesis is as much yours as it is mine!

Finally, I would like to dedicate this thesis to my Grandad Tom- Your love of acquiring knowledge instilled in me a passion for learning and development and although I am not eligible for a bus pass just yet, (maybe I’ll have to do some more study 😘), I hope you’re still finding enjoyment in this joke, Jennifrog 😊
Publications and Conference Presentations Resulting from this Thesis

Publications


Manuscripts in Preparation


Conference Presentations


Poster Presentations


List of Figures and Tables

Figures

Chapter 2 Behaviour Analytic Approaches to Language
Figure 2.1. Examples of non-arbitrary and arbitrary relational responding...........25

Chapter 4 The Relational Evaluation Procedure (REP)
Figure 4.1. Examples of REP trials.................................................................56
Figure 4.2. A trial from the final stage of Stewart et al. (2004) which involved using an REP methodology to train and test participants in analogical reasoning .................................................................60
Figure 4.3. Example of a trial on the SMART online training program..............62
Figure 4.4. The four IRAP trial types ..............................................................65
Figure 4.5 Illustrations of each of the three levels appearing in the NSD-REP.....72

Chapter 5 Assessment and training of Yes/No Question (Y/N-Qs) responding in children
Figure 5.1. Analysis of the Joint Control process with Y/N-Qs .........................84

Chapter 6 Assessing and Training Relational Evaluation in Young Children
Figure 6.1. Examples of the three levels in the NSD-REP ..............................95-96
Figure 6.2. Study 2. Percentage of correct responses for Level 1 of the NSD-REP during PI baseline (BL), PII intervention (INT), PIII baseline (BL) and PIV follow-up (FU), across each participant .........................................................108
Figure 6.3 Study 3. Percentage of correct responses for level 2 of the NSD-REP during phase I baseline (BL), phase II intervention (INT), phase III baseline (BL) and phase IV follow-up (FU), for each of the eight participants involved ...........................................................................115
Figure 6.4 Study 4. Upper panel: Percentage of correct responses for Level 1 of the NSD-REP during various phases (see legend) for participant 4.1. Lower panel: Percentage of correct responses for level 1 of the NSD-REP during various phases (see legend) for participant 4.2...............120-121
Chapter 7 Assessing and Training Relational Evaluation in Children with Autism Spectrum Disorder

Figure 7.1 Study 6. Percentage of correct responses for Level 2 of the NSD-REP during phase I baseline (BL) and phase II intervention (INT) for each of the three participants involved ................................. 146

Chapter 8 Assessing and Training Relational Evaluation in Mid-Childhood using the NSD-REP

Figure 8.1. Example of a Relational Abilities Index (RAI) trial.......................... 166
Figure 8.2. Study 8. Percentage of correct responses for Level 3 of the NSD-REP during phase I baseline (BL), phase II intervention (INT), phase III baseline (BL) and phase IV follow-up (FU), for each of the three participants involved ................................................................. 180
Figure 8.3. Study 9. Data for participants (Pts 1-3) in multiple baseline design with NSD-REP Level 3 (not same/not different) responding (Phases 1,5 and 6) as the target and with multi-phase (2-4) intervention training.............. 190

Chapter 9 Comparing the effects of derived relational training and computer coding on intellectual potential in school age children

Figure 9.1. Example of the Scratch computer coding interface.......................... 209
Figure 9.2. Graphs comparing mean performance for the Experimental and Control groups at both pre- and post-testing on each of the dependent measures ................................................................. 212
Figure 9.3. Graph comparing mean WASI Full-Scale IQ scores for the Experimental and Control groups at both pre- and post-testing ......................... 215

Tables

Chapter 6 Assessing and Training Relational Evaluation in Young Children

Table 6.1 Demographic Information, Scores for the NSD-REP, PLS4 and SB5-ABIQ assessments ................................................................. 99-100
Table 6.2 Spearman’s rank order correlations (n = 26) between key assessment measures ................................................................. 101
Table 6.3 Demographics and scores at pre and post-test for SB5-ABIQ, PLS4 and NSD-REP ..........................127-128

Chapter 7 Assessing and Training Relational Evaluation in Children with Autism Spectrum Disorder

Table 7.1. Demographic Information, Pre-Intervention NSD-REP and Pre-Intervention PLS4 Scores ..........................................................144
Table 7.2. Ages, scores on NSD-REP and score and age equivalence for PLS4 at pre- and post-intervention ..........................................................148

Chapter 8 Assessing and Training Relational Evaluation in Mid-Childhood using the NSD-REP

Table 8.1. Demographic Information, Scores for NSD-REP, RAI SB5 & Drumcondra Assessments ...............................................................168-169
Table 8.2 Spearman’s rank order correlations (n = 23) between key assessment measures .................................................................171
Table 8.3 Baseline and Intervention Scores for Participants 1-3 using unadjusted NSD-REP Level 3 .................................................................184

Chapter 9 Comparing the effects of derived relational training and computer coding on intellectual potential in school age children

Table 9.1. Number of weekly sessions for participants, number of sessions missed, additional sessions in school and home and Level attained in the SMART condition .................................................................206-207
Table 9.2. Mean performance for the experimental and control group on each dependent measure (raw scores) at pre and post-test ...................211
Table 9.3. Correlation (Pearson’s r) of RAI & Other Measures at Pre-intervention .................................................................216
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>REP</td>
<td>Relational Evaluation Procedure</td>
</tr>
<tr>
<td>ASD</td>
<td>Autism Spectrum Disorder</td>
</tr>
<tr>
<td>DRR</td>
<td>Derived Relational Responding</td>
</tr>
<tr>
<td>SE</td>
<td>Stimulus Equivalence</td>
</tr>
<tr>
<td>JSC</td>
<td>Joint Stimulus Control</td>
</tr>
<tr>
<td>RFT</td>
<td>Relational Frame Theory</td>
</tr>
<tr>
<td>AARR</td>
<td>Arbitrary Applicable Relational Responding</td>
</tr>
<tr>
<td>MET</td>
<td>Multiple Exemplar Training</td>
</tr>
<tr>
<td>MSR</td>
<td>Multiple Stimulus Relations</td>
</tr>
<tr>
<td>MTS</td>
<td>Match-to-Sample</td>
</tr>
<tr>
<td>OTM</td>
<td>One-to-Many</td>
</tr>
<tr>
<td>MTO</td>
<td>Many-to-One</td>
</tr>
<tr>
<td>SMART</td>
<td>Strengthening Mental Abilities with Relational Training</td>
</tr>
<tr>
<td>IRAP</td>
<td>Implicit Relational Assessment Procedure</td>
</tr>
<tr>
<td>RCP</td>
<td>Relational Completion Procedure</td>
</tr>
<tr>
<td>RCI</td>
<td>Relational Coherence Indicator</td>
</tr>
<tr>
<td>Y/N-Qs</td>
<td>Yes/No Questions</td>
</tr>
<tr>
<td>MBD</td>
<td>Multiple Baseline Design</td>
</tr>
</tbody>
</table>
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS4:</td>
<td>Preschool language scales - 4th Edition</td>
</tr>
<tr>
<td>SB5:</td>
<td>Stanford Binet Intelligence Scales - 5th Edition</td>
</tr>
<tr>
<td>ABIQ:</td>
<td>Abbreviated Battery IQ</td>
</tr>
<tr>
<td>RAI:</td>
<td>Relational Abilities Index</td>
</tr>
<tr>
<td>DPRT-R:</td>
<td>Drumcondra Primary Reading Test - Revised</td>
</tr>
<tr>
<td>DPMT-R:</td>
<td>Drumcondra Primary Math Test - Revised</td>
</tr>
<tr>
<td>WASI:</td>
<td>Wechsler Abbreviated Scale of Intelligence</td>
</tr>
<tr>
<td>WISC-IV:</td>
<td>Wechsler Intelligence Scale for Children-4th Edition</td>
</tr>
<tr>
<td>WIAT-II:</td>
<td>Wechsler Individual Achievement Test – 2nd Edition</td>
</tr>
</tbody>
</table>
The principal goal of this thesis is to explore the utility of the Relational Evaluation Procedure (REP; e.g., Stewart, Barnes-Holmes, & Roche, 2004) in assessing and training relational skills in children, both neurotypical, as well as developmentally delayed. The REP protocol is derived from Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), a contextual behaviour analytic approach to language and cognition which has seen significant empirical progress over the last few decades.

As will be explicated in Chapter 2, RFT explains language and complex cognitive phenomena in terms of the operant of derived relational responding (DRR). Hence, understanding and exploring DRR is of core importance to understanding human psychology, from an RFT perspective. Up until now, the predominant means of studying DRR within behaviour analytic science has been the matching-to-sample (MTS) procedure. MTS facilitates the study of DRR by enabling the training and assessment of conditional discriminations. That is, under certain conditions, trained patterns of conditional discriminations are seen to give rise to the emergence of further untrained or derived performances; hence the name derived relational responding. However, the dominance of MTS as a methodology for studying derived stimulus relations, from an RFT viewpoint, is not conducive to the most effective progress of research into this phenomenon. The primary reasons for this include: (i) MTS is relatively slow and inefficient as a means of exploring this phenomenon and; (ii) it is limited with respect to the detection of the full range of types of derived stimulus relations. RFT researchers (e.g., Hayes and Barnes, 1997) have argued for the need to develop new methodologies so that DRR might be analysed and experimentally manipulated efficiently and across its full range. Consequently, the REP was developed.
As with other key phenomena touched on here in this opening chapter, much more detail will be provided on the REP in later chapters. For now, it is sufficient to say that this protocol facilitates exploration of relational responding by inducing participants to “evaluate”, or “report on”, stimulus relations that are presented. In a typical approach, participants may confirm or deny the applicability of particular stimulus relations to other sets of stimulus relations (Stewart et al., 2004). Recently, the REP has also evolved into several forms of protocol (e.g., the Strengthening Mental Abilities with Relational Training [SMART]; the Implicit Relational Assessment Procedure [IRAP]; and the Relational Completion Procedure [RCP]), that are useful for research into DRR, and hence, the full range of complex human language and cognition.

Despite its potential, to date, most work using the REP has been in the assessment and training of relational framing in adults (Stewart et al., 2004; Roddy, Stewart, & Barnes-Holmes, 2010). It has been much less prevalent in empirical investigations with children, though a small number of studies with older neurotypical children and children with developmental delay have started to emerge (e.g., Cassidy, Roche, & Hayes, 2011; Kilroe, Murphy, Barnes-Holmes, & Barnes-Holmes, 2014). Given the importance of DRR for understanding human psychology and the potential of the REP methodology, more research investigating the utility of this protocol, in children specifically, is needed. This is a key aim of the current thesis. To this end, across ten empirical studies, the efficacy of the REP, as an assessment and training protocol in the context of children, both neurotypical as well as developmentally delayed, is examined. In what follows, the information presented across the chapters of the research thesis is summarised.
Chapter 2 presents the relevant background theory for this research by detailing behaviour analytic approaches to language. To start with, the Skinnerian contribution is discussed, including Skinner’s 1957 operant account of language development and its successful application within the field of behavioural intervention. Following this, the conceptual and applied limitations of his theory are highlighted, most notably, the inadequacies with respect to explaining linguistic generativity (i.e., the ability to produce and understand novel sentences) in human language (e.g., D. Barnes-Holmes, Y. Barnes-Holmes, McHugh, & Hayes, 2004; Hayes et al., 2001). More contemporary behavioural theories of language and cognition, including Stimulus Equivalence (Sidman, 1971), Naming Theory (Horne & Lowe, 1996), and Joint Stimulus Control (Lowenkron, 1998) are subsequently detailed. This is followed by comprehensive delineation of RFT, an approach which suggests that DRR is the key process involved in generative language. The successful application of this theory, with specific reference to typically developing children and children with developmental delay, is then discussed in Chapter 3.

Towards the end of Chapter 3, the REP is introduced as an efficient RFT-based methodology for assessing and training relational framing. However, a more in-depth discussion is left to Chapter 4. The latter describes the REP and some of its major research outgrowths (e.g., SMART, IRAP and RCP). As indicated above, one of the features of REP research is that it has predominantly centred on adult participants. Considering the potential of the REP in the assessment and in particular the training of DRR, much more REP work needs to be conducted with children. One key aspect of starting such work, is to examine children’s capacity to be effectively assessed and trained using an REP like methodology. This exploration is the primary focus of the current thesis.
Responding with yes/no or alternate affirmation/disconfirmation relations (e.g., true/false, same/different), is critical in the REP. Thus, Chapter 5 presents an overview of the development and training of yes/no responding in children. The ability to respond correctly using a yes or no response is emphasized as a basic but fundamental language repertoire that plays a pivotal role in social interactions (e.g., Raymond, 2003). A number of research areas concerning the development and training of this fundamental language skill are subsequently outlined. These include: (i) the frequency of yes/no questions with children used within the cognitive and developmental literature; (ii) the outcomes of studies exploring patterns of responding to yes and no questions among young children, specifically with respect to patterns of bias shown at different ages and from different cultural backgrounds; and (iii) the employment and exploration of yes and no questions in young children with developmental delay. Following the presentation of this background information, the studies forming the current line of research are then described.

The empirical investigations conducted as part of this thesis are laid out in chapters six through nine. All of these studies, with the exception of Study 10, utilise the NSD-REP (Non-arbitrary Same and Different Relational Evaluation Procedure), a multi-level assessment and training instrument. Typically, research using the REP, would focus on arbitrary (abstract) relations and indeed, arbitrary relations are essential for the demonstration of DRR. The NSD-REP, in contrast, focuses on non-arbitrary (physical) rather than arbitrary relations, requiring participants to evaluate these correctly under the control of contextual cues (e.g., “Same” and “Different”, “Yes” and “No”). The use of this protocol is adopted in order to facilitate the work conducted with young children so as not to confound inability to come under appropriate REP control with lack of familiarity with
particular arbitrary relations. This point will be returned to later on (i.e., in Chapter 6).

In the studies in Chapter 6, the NSD-REP is employed as both an assessment (n = 26) and training (n = 12) protocol in very young (i.e., aged 2-4 years), typically developing children. Study 1 correlates performance on the NSD-REP protocol with performance on standardised measures of language and cognition, specifically the Preschool Language Scale 4th Edition (PLS4; Zimmerman, Steiner, & Pond, 2002) and the Stanford Binet Intelligence Scales 5th Edition (SB5; Roid, 2003). The results of this study are described and discussed. The second (n = 4), third (n = 8) and fourth (n = 2) studies explore the use of the NSD-REP as a training protocol in very young children. Using both concurrent and non-concurrent multiple baseline designs (MBDs) across participants, the NSD-REP is used to teach children yes/no responding to simple non-arbitrary same and different relations across two levels (i.e., Levels 1 and 2). Finally, in Study 5, the data from Studies 2-4 is analysed as a means of providing a preliminary test of the effects of NSD-REP training on the intellectual (linguistic and cognitive) potential of the participants involved in these studies.

Study 6 (Chapter 7) extends the work presented above, by exploring the efficacy of the NSD-REP as an assessment (n = 9) and training tool (n = 3) in children with Autism Spectrum Disorder (ASD). Initially, performance on the NSD-REP is correlated with performance on the PLS4. Subsequently, using a concurrent MBD across participants, three children, aged 7-9 years, are taught to respond to the second level of the NSD-REP, involving the affirmation and disconfirmation of non-arbitrary same and different relations. To conclude Study 6,
PLS4 scores for the three participants who were taught to successfully respond on the NSD-REP are re-examined.

Chapter 8 presents the next series of studies (i.e., Studies 7-9), which examine the use of the NSD-REP as an assessment and training protocol in slightly older typically developing children (i.e., 6-7 years). This allows further verification of the NSD-REP as a potential assessment and training tool, as well as providing some measure of comparison of performance across differing age groups. Study 7 explores the relationship between the NSD-REP, the SB5, the Relational Abilities Index (RAI), and standardised school assessments, specifically the Drumcondra Primary Reading Test- Revised (DPRT-R) and the Drumcondra Primary Math Test- Revised (DPMT-R). The quantitative correlations between the assessments are analysed and discussed. Studies 8 (n = 3) and 9 (n = 3) follow and advance this work further by training children on Level 3 of the NSD-REP, which involves negation controlled responding. Utilising concurrent MBDs, older children (6-7 years) are trained simply through feedback, while younger children (around 4 years) require additional sub-task training to facilitate performance.

Study 10 (Chapter 9) reports work with the oldest group of children in the current programme, namely 10-12 year olds. This study (n = 28) utilises a more sophisticated version of the REP (i.e., the SMART), which includes a greater range of relations, suitable for use with this age range. It also extends previously published research (i.e. Cassidy et al., 2011) by exploring the effect of derived relations training under more controlled conditions. Using a group-based design, standardized tests of IQ and academic attainment are compared at pre and post-test for two cohorts of participants, those in the SMART group (who received training with same/opposite, more/less relations) and those in the control condition (i.e., Scratch –
a computer coding program). The results of this final study are outlined and analysed.

This thesis then concludes with Chapter 10, the General Discussion. A recap of the primary research aims, as well as the individual aims and the findings of the various empirical studies are outlined. Subsequently, the theoretical issues and implications of the empirical work (including possible limitations as well as potential directions for future research) are discussed.
Chapter 2

Behaviour Analytic Approaches to Language
The theory that underlies the Relational Evaluation Procedure (REP) is Relational Frame Theory (RFT: Hayes, Barnes-Holmes, & Roche, 2001), a behaviour analytic account of language and cognition. In order to provide theoretical background to the work presented herein, this chapter commences by examining behavioural approaches to language and generative behaviour that pre-date RFT, along with their conceptual and applied limitations. Subsequently, RFT will be introduced as an alternative, offering several important advantages over previous accounts. The core theory of RFT will be described, along with the properties of relational framing.

**Language Within Behaviour Analysis**

Language is a ubiquitous and profoundly influential aspect of human behaviour. It is of fundamental importance in life and society, including key domains of activity such as communication, education, problem-solving, culture and science. For this reason, scholars across academic disciplines have sought to better understand the environmental, biological and neural factors that underpin this ability, and to identify its core properties, structure and function.

Mainstream psychologists typically conceptualize language mentalistically, as being mediated by a specific set of mental or computational processes which facilitate linguistic comprehension and production. This has resulted in an emphasis on the structural properties of language (e.g., morphology, syntax, and phonology) as well as the mental mechanisms and knowledge representations necessary for its development and operation (for a more detailed treatment see Harley, 2014).

In contrast, from a behaviour analytic perspective, language is not considered an innate, mentalistic process. Rather, language is conceptualised as an operant behaviour that it is acquired, extended, and maintained by environmental variables.
As such, the term “Verbal Behaviour” has been used as an alternative to language to suggest its environmental determination. Within the behaviour analytic tradition, several accounts of language acquisition and development, have been proposed, the first of which was provided by B.F. Skinner.

**Verbal Behaviour**

Skinner’s (1957) book *Verbal Behavior* was a landmark theoretical work. It was the first treatment of language as operant behaviour, described in the same terms as all other operant behaviour: learned through a history of reinforcement, and influenced by environmental variables. In this book, Skinner defined verbal behaviour as “behavior reinforced through the mediation of other persons” (Skinner, 1957, p.2). This definition was further developed by proposing that “[v]erbal behavior is shaped and sustained by a verbal environment — by people who respond to behavior in certain ways because of the practices of the group of which they are members” (p.226).

Emphasizing response functionality over topography, Skinner’s analysis of verbal behaviour described several specific classes of verbal operants that comprise a language repertoire. These include the echoic, mand, tact, intraverbal, autoclitic, transcription and textual behaviour. In order to highlight the functional nature of Skinner’s analysis, a brief description and example of some of the verbal operants, will be provided, prior to addressing the utility of this approach as a theory of language.

The *echoic* is an operant involving a response that is evoked by a verbal discriminative stimulus ($S_D^D$) which has point-to-point correspondence and formal similarity with the response. As illustration, when a parent says “juice”, the child says “juice” in response. The importance of an echoic repertoire is highlighted by
the fact that it may facilitate the acquisition of other verbal operants (Esch, Mahoney, Kestner, LaLonde, & Esch, 2013). As such, research utilising echoic responding has primarily focused on the use of echoic behaviour in stimulus control procedures for establishing other verbal operants (e.g., Kodak, Fuchtman, & Paden, 2012).

The mand is an operant involving a response that is evoked by a Motivating Operation (MO) and followed by specific reinforcement. For example, when a young child wants juice, the child may request it from a parent by saying “juice?” The parent in turn delivers the reinforcer specific to the mand topography, in this case, juice. The mand repertoire is essential for early language learners (Albert, Carbone, Murray, Hagerty, & Sweeney-Kerwin, 2012). In addition, manding helps establish the reciprocal speaker and listener roles that are essential for increasing verbal competence (Sundberg & Michael, 2001). As such, a considerable amount of the empirical literature, from a Skinnerian verbal behaviour perspective, has focused on the mand (Sautter & LeBlanc, 2006; Dixon, Small, & Rosales, 2007). Applied research has included studies ranging from the development of functional communication training procedures (e.g., Carr & Durand, 1985) to procedures for teaching initial vocal mands to young children with autism (Ross & Greer, 2003).

The tact is an operant involving a response that is evoked by a nonverbal S<sup>D</sup> and followed by generalized conditioned reinforcement. For instance, a child may see, point to and vocally respond “juice” (tact) upon seeing the actual juice (nonverbal S<sup>D</sup>), to which the parent responds “That is juice!” (generalized conditioned reinforcement). Skinner (1957) described the tact as the most important of the verbal operants because of the ability of listeners in the natural environment to provide nonspecific reinforcement that can function as an S<sup>D</sup> for additional verbal
behaviour. Research on tacting behaviour encompasses procedures for increasing the variability of tacts (e.g., Heldt & Schlinger, 2012), the development of transfer of stimulus control procedures for effectively teaching tacts (e.g., Sundberg, Endicott & Eigenheer, 2000), and the appropriate sequence of teaching tacts and listener responses (see Petursdottir & Carr, 2011).

Intraverbals are operants involving a response that is evoked by a verbal SD that does not have point-to-point correspondence with that verbal stimulus. In other words, parts or subdivisions of the stimulus do not directly control parts or subdivisions of the response (e.g., saying “juice” as a response to the question “what would you like to drink?”). Intraverbal behaviour constitutes the basis for social interaction, conversations, and much of academic behaviour (Partington & Bailey, 1993). Although less prevalent than research on alternative verbal operants (i.e., mand, tact and echoic), empirical strategies for establishing or expanding intraverbal responding include: (i) the transfer of stimulus control (e.g., Braam & Poling, 1983; Luciano, 1986); (ii) peer mediated interventions (e.g., Krantz, Ramsland, & McClannahan, 1989); (iii) video modeling (e.g., Sherer Pierce, Paredes, Kisacky, Ingersoll, & Schreibman, 2001); (iv) conversation skills training programs (e.g., Lewis, Roessler, Greenwood, & Evans, 1985); (v) discrete trial training (e.g., Wong and Woolsey, 1989); and (vi) direct instruction (e.g., Binder & Watkins, 1990). However, despite the complexity of this repertoire and utility of these methodologies, the vast majority of empirical work has concentrated on appraising teaching methodologies for relatively simple intraverbal responses (e.g., Ingvarsson & Hollabough, 2011; Kodak et al., 2012).
Applications and Limitations of Skinner’s *Verbal Behavior*

Skinner’s *Verbal Behavior* was the first comprehensive behavioural account of language. Though his research primarily involved the laboratory study of animals, the publication of this book instigated the experimental analysis of verbal behaviour in humans. Bijou’s early efforts (e.g., Bijou, 1955, 1957, 1958), in which he detailed specific instrumentation, how data was to be recorded, and how to maximize control over independent variables, were fundamental to this (Dixon, Vogel and Tarbox, 2012). His theory of child language and development also served as the theoretical basis for the application of behavioural principles to the education of children.

Hence, extensions of Skinner’s theory have had utility in developing procedures for targeting language skills in individuals with developmental disabilities, most notably Autism Spectrum Disorder (ASD) (e.g., Greer & Ross, 2008; Schauffler & Greer, 2006; Stafford, Sundberg, & Braam, 1988; Sundberg & Michael, 2001; Sundberg & Partington, 1998). For instance, the empirical base for teaching various verbal operants to children with ASD has grown over several decades (Sautter & LeBlanc, 2006). There are many experimental studies on components of the *Verbal Behavior* approach (Cautilli, 2007; Prelock, Paul, & Allen, 2011; Sundberg & Michael, 2001) and research is beginning to emerge regarding the efficacy of full *Verbal Behavior* programmes for students with ASD (e.g., Bondy, Esch, Esch, & Sundberg, 2010; Miklos, Dipuglia, & Galbraith, 2010).

New behaviour analytic research has also started to emerge regarding the investigation of verbal operants in neurotypical children (e.g., Haq & Kodak, 2015; Pétursdóttir, Carr, Lechago, & Almason, 2008; Sundberg & Sundberg, 2011). For instance, Haq and Kodak (2015) compared the efficacy of two alternative
methodologies (i.e., massed and distributed practice) on the acquisition of textual and tacting behaviour in two neurotypical children aged four and ten years. Both participants were exposed to two sets of target stimuli. For the younger participants this included: (i) English sight words; and (ii) pictures of common nouns (e.g., bed, turkey), for which a tact in Spanish was required. For the older participant, both sets of target stimuli contained verbs printed in English, to which a textual response in Spanish was required. Subsequent to training, the researchers reported that although both procedures were successful, the distributed practice condition was a more efficient training format.

The results from the aforementioned investigations illustrate that Skinner’s analysis has had utility in teaching language skills to children with and without disabilities. Yet, while his work has had a substantial influence on the basic conceptual framework of behaviour analysis, it has been argued that at a theoretical and applied level, *Verbal Behavior* is deficient as an approach to language (Hayes et al., 2001).

Skinner’s theory of language has been criticised by both linguists and mainstream psychologists. For instance, Chomsky (1959) asserted that language must be mediated by a mental component as it is not conceivable for a child to acquire such an extensive repertoire by direct instruction alone. He contended that Skinner’s account failed, amongst other things, to provide a satisfactory explanation for linguistic generativity, the ability to produce and understand novel linguistic constructions (a detailed description of this is provided in the next section).

Although Chomsky’s review was seen as erroneous by many within the behaviour analytic community (e.g., MacCorquodale, 1970; Schlinger, 2010), it was widely accepted by mainstream psychologists. This contributed, in part, to the historical
shift towards research interested in the mental mechanics of language. Specifically, mainstream psychologists switched their focus away from functional analyses of behaviour-environment interactions, to the mental level of analysis and began postulating hypothetical or “computational” mechanisms to explain how language was acquired and utilised.

Yet, criticism of *Verbal Behavior* has also come from within the behaviour analytic field. Researchers have argued that Skinner’s analysis has been difficult to turn into a progressive, empirical research programme (Hayes et al., 2001). Further, it has been disseminated as a complete package mostly via workshops, non-peer reviewed articles (Barbera, 2009a), conference presentations (Barbera, 2009b; Carbone, 2004; Miklos et al., 2010), and books (Barbera & Rasmussen, 2007; Schramm, 2011; Sundberg & Partington, 1998). Though there is a modest amount of research supporting the components of this approach, without empirical validation of the package as a whole, Carr and Firth (2005) have suggested that the widespread dissemination of *Verbal Behavior* occurred too prematurely.

Several reviews have also examined the impact that *Verbal Behavior* has had on empirical research. The results indicate, that overall, the effects have been limited. McPherson, Bonem, Green, and Osborne (1984) reviewed studies published between 1961 and 1980 for citations of the term *Verbal Behavior* and reported that while over 800 studies cited the term, only 31 of these studies explicitly examined one of Skinner’s verbal operants. In a more recent investigation, Sautter and LeBlanc (2006) found that most of the published empirical research on Skinner’s analysis of verbal behaviour has involved the development, assessment, or analysis of a mand repertoire either individually or in conjunction with other verbal operants. The results of this review emphasize, that while there is evidence supporting
Skinner’s conceptualization and taxonomy, many areas of verbal behaviour research have yet to be addressed. Some of the reasons cited in this quantitative review included: (i) a problematic definition; and (ii) failure to capture the essence of conversation, namely generative language.

In support of this, Hayes et al. (2001) suggest that the definition utilised by Skinner is not functional and is too broad. It is not functional because it is not based on aspects of an individual organism’s history but by an audience trained to mediate reinforcement to the speaker. Further, it is so broad that it includes practically all animal operant behaviour in traditional behaviour analytic research. In addition, Skinner’s analysis focused mainly on the composition of simple, independent elements of verbal behaviour. Yet it remains unknown whether these concepts promote a rubric for analyzing and influencing conversation. While discourses may include numerous elemental verbal operants such as mands or intraverbals, it remains possible that these operants do not define or capture the essence of conversations or facilitate the type of generativity characterising typical language.

So far, this chapter has outlined Skinner’s account of language acquisition and development, as well as the applications and limitations of this approach, most notably with respect to linguistic generativity. In the next sections, derived relational responding and its relevance to understanding generativity in language will be detailed.

**Linguistic Generativity and Derived Stimulus Relations**

Linguistic generativity may be described as the ability to produce or understand novel sentences (Hayes et al., 2001). It is imperative to the development of fully functional communication and thus, the development and/or training of this phenomenon is critical. Much of the criticism of Skinner’s operant approach within
behaviour analysis has come from proponents of Relational Frame Theory (RFT). These theorists partly agree with Chomsky’s criticism of Skinner’s account in terms of its inadequacy with respect to explaining the generativity of human language (e.g., Barnes-Holmes, Barnes-Holmes, McHugh, & Hayes, 2004; Hayes et al., 2001).

Generative responding is seen as a significant indicator of progress in the behaviour analytic literature. It is often discussed in terms of response generalisation (e.g., Stewart, McElwee & Ming, 2013). Yet, as a concept, response generalisation, is problematic for two reasons. First, an adequate history of learning responsible for response generalisation that can be easily adapted to educational programming has not been detailed. Second, the parameters in which response generalisation has been described have been poorly defined. Some theorists speak of functionally equivalent behaviours (e.g., Cooper, Heron & Heward, 2007) while others incorporate physical dimensions into their definitions (e.g., Mayer, Sulzer-Azaroff & Wallace, 2011).

Alternative behaviour analytic research conducted since Skinner has uncovered a phenomenon referred to as derived relational responding (DRR) which seems extremely relevant with regards understanding generativity in language. Moreover, it has provided the basis for considerable empirical work and theoretical development. The capacity to derive relations is a learned ability that is fundamental in the facilitation of human language. In what follows, a theoretical background to DRR and RFT is provided.

**Stimulus Equivalence**

The first example of DRR was provided by Sidman (1971). Sidman showed that by training a series of related conditional matching performances using non-arbitrary stimuli in a young man with a learning disability, several untaught performances emerged. As illustration, prior to the commencement of the
experiment, the participant had the ability to select particular pictures (X) in the presence of corresponding spoken words (Y) and could produce the appropriate spoken words (Y) in the presence of the pictures (X). During training, he was then taught to pick appropriate textual stimuli (Z) in the presence of the corresponding spoken words (Y). Subsequent to the training phase, the participant then exhibited various derived or untaught performances. For instance, he was capable of producing the appropriate spoken words in the presence of textual stimuli which was a reversal of the taught performance (i.e., Z \(\rightarrow\) Y). In addition, the participant was also able to select the appropriate textual stimuli in the presence of pictures and vice versa (i.e., Z \(\rightarrow\) X and X \(\rightarrow\) Z).

Sidman named the overall pattern stimulus equivalence because it seemed as though the participant was suddenly treating the stimuli involved as mutually substitutable or equivalent (i.e., the participant was responding as if particular sets of spoken words, pictures and printed words were the same as or equivalent to each other). Based on these and later results (1971, 1980, 1986), Sidman suggested that stimuli are members of an equivalence class when conditional discrimination performance shows the properties of reflexivity, symmetry, and transitivity. For instance, reflexivity occurs when each stimulus is matched to itself (e.g., X-X, Y-Y, Z-Z); symmetry occurs when each of the trained relations is reversed (e.g., Y-X, Z-Y); and transitivity occurs when the two (or more) separate trained relations combine to produce a novel relation (e.g., X-Z, Z-X).

The phenomenon of stimulus equivalence has generated a great deal of interest within behaviour analytic research because it is not predicted by operant conditioning (Barnes, 1994) and it is efficient (i.e., not all relations need be taught directly). However, perhaps the most compelling reason to study this phenomenon
was that empirical research suggested that there was a strong link with language (Cowley, Green, & Braunling-McMorrow, 1992; Devany, Hayes, & Nelson, 1986; Kendall, 1983; Wulfert & Hayes, 1988). Since the 1980s, the study of derived equivalence responding has been directly linked to the behaviour analysis of human language in a variety of different contexts (Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan, & Leader, 2004).

For example, Barnes (1994) outlined five areas of research that provide empirical evidence in support of the view that equivalence and human language are closely related. First, whereas derived equivalence is readily demonstrated by verbal humans, it has not been unequivocally demonstrated by non-humans or by humans who have limited verbal repertoires (Barnes, McCullagh, & Keenan, 1990; Devany et al., 1986; Dugdale & Lowe, 2000; Hayes, 1989; Sidman & Tailby, 1982).

Second, learning to name stimuli may facilitate equivalence responding in young children (Dugdale & Lowe, 1990). Third, equivalence procedures should prove useful in teaching basic verbal skills such as reading, to the verbally impaired (e.g., Cowley et al., 1992; Sidman & Tailby, 1982). Fourth, stimulus equivalence has been used to develop a behaviour-analytic interpretation of both symbolic meaning and the generative nature of grammar (Barnes & Holmes, 1991; Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000; Hayes & Hayes, 1989; Wulfert & Hayes, 1988). Fifth, equivalence phenomena have been used to examine verbal human behaviours such as social categorisation (e.g., Roche & Barnes, 1996; Watt, Keenan, Barnes, & Cairns, 1991) and logical reasoning (Barnes & Hampson, 1993).

In support of the empirical evidence outlined by Barnes (1994), the link between DRR and language is also strengthened by the results of neuroimaging studies. These illustrate that brain activity measured during derived equivalence
Theories of Language

tasks resembles that seen during language performance (e.g., Ogawa, Yamazaki, Ueno, Cheng, & Iriki, 2010). Overall, there is substantive evidence that the control exerted over behaviour by stimuli participating in equivalence classes, parallels the control that verbal stimuli exert over human behaviour (Hayes & Hayes, 1989). Also pertinent in this respect, is the related phenomenon of transfer of function.

Transfer of Function

When a particular behavioural function is established for one of the stimuli in an equivalence relation and the function transfers to the remaining class members without further training, it is referred to in the literature as the “transfer of function” effect. (e.g., Dougher, Auguston, Markham, Greenway, & Wulfert, 1994). This phenomenon has been demonstrated in numerous research studies. For example, in Barnes, Browne, Smeets, and Roche (1995), children were initially trained and tested for the following equivalence relations (X1-Y1-Z1, X2-Y2-Z2). They were subsequently trained to perform specific actions in the presence of particular stimuli involved in those equivalence relations (e.g., clapping in the presence of Z1 and waving in the presence of Z2). Following this, the participants illustrated these actions in the presence of stimuli in derived relations with the original ones (i.e., clapping in the presence of X1 and waving in the presence of X2). This empirically demonstrated phenomenon, which has been suggested to model the process of linguistic control, is tremendously efficient in relation to training novel repertoires and thus is of particular practical benefit.

Derived Relational Responding (DRR)

As aforementioned, stimulus equivalence (and the accompanying transfer of function effect), is closely related to language and is extremely generative. Moreover, stimulus equivalence is also the most researched example of DRR. Yet, it
is not the only form of derived relation for which there is empirical evidence. Several other patterns of derived relations have been detailed in the literature including, for example, comparison (e.g., Dougher, Hamilton, Fink, & Harrington, 2007), distinction (e.g., Steele & Hayes, 1991), opposition (e.g., Dymond, Roche, Forsyth, Whelan, & Rhoden, 2007), deixis (e.g., McHugh, Barnes-Holmes, Barnes-Holmes, & Stewart, 2006), analogy (e.g., Stewart, Barnes-Holmes, Roche, & Smeets, 2001) and temporality (e.g., O'Hora, Barnes-Holmes, Roche, & Smeets, 2004).

Given the importance of the relationship between DRR and language, one question that arises is how to explicate this phenomenon 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 et al., 2001).

Sidman (1994; 2000) has argued that DRR is a phenomenon that is explained through reinforcing contingencies, probably due to phylogenesis. This means that DRR is a basic behavioural process just like reinforcement, generalisation and discrimination except that equivalence is a faculty 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). In addition, although the study of derived equivalence responding has been directly linked to the behaviour analysis of human language in a variety of different contexts (e.g., Barnes & Hampson, 1993; Dugdale & Lowe, 1990, 2000; Roche & Barnes, 1996; Sidman & Tailby, 1982; Wulfert & Hayes, 1988), it has also introduced theoretical difficulties. An empirical relationship does not indicate that equivalence relations depend upon, or are mediated by language, nor does it indicate
that language depends upon equivalence relations (see Carr, Wilkinson, Blackman, & McIlvane, 2000 and Kastak & Schusterman, 2002). In addition, Sidman’s theory does not appear to provide a generic nomenclature for the complex patterns inherent in non-equivalence relations. For example, take the relation of “smaller than”. If X is smaller than Y, it does not follow that Y is also smaller than X. The relation between the stimuli is not symmetrical, and thus the stimuli do not meet the criteria necessary to be categorised as participating in an equivalence class. Put simply, there are other forms of DRR (e.g., distinction, temporal) for which Sidman’s theory cannot account and these alternative derived stimulus relations have also been associated with language performance (e.g., O Hora, Pelaez & Barnes-Holmes, 2005).

A number of behavioural accounts other than Stimulus Equivalence have been put forward in an attempt to provide an explanation of DRR and its empirical link with language. These include RFT, Naming Theory and Joint Stimulus Control (JSC) (e.g., Hayes et al., 2001; Horne & Lowe, 1996; Lowenkron, 1998). Relating to RFT specifically, DRR is explained as generalized contextually controlled patterns of responding (see Barnes, 1994; Stewart & McElwee, 2009). These are based on a history of multiple exemplar training (MET) in which the functions of the contextual cues controlling the patterns involved, are established. The approach taken by RFT will be explicated in the next section. Prior to this, however, Naming and JSC will be briefly described (a more detailed account of JSC will be provided in Chapter 5, pp. 83-84).

These accounts are comparable in that they explain DRR as being based on a covert process (either naming or joint control, respectively) that is established through multiple exemplar training (MET), which subsequently enables DRR to
occur. From a behavioural perspective, a fundamental theoretical criticism of both accounts is that they reference a covert mediational process, which is inconsistent with the radical behavioural goal of gaining prediction and influence over behaviour. Further, reliance on these additional processes makes these theories less parsimonious than RFT, which does not (Stewart et al., 2013). In addition, there is also empirical evidence refuting Naming. For instance, derived relations can emerge in the absence of naming (see Luciano, Gómez & Rodríguez, 2007 as illustration). Finally, there is little or no elaboration of the naming or joint stimulus control account with respect to types of derived relations other than equivalence and as suggested, these have also been correlated with language performance.

This chapter has now outlined several theories that have attempted to detail the empirical relationship between DRR and language at a conceptual level. These include Stimulus Equivalence, Naming Theory and Joint Stimulus Control, the limitations of which have been described in the preceding paragraphs. RFT (Barnes-Holmes et al., 2001; Hayes et al., 2001; Hayes, 1991; 1992) is the only residual behaviour analytic theory that endeavours to provide an explanation of DRR and its empirical association with language. As the work in current research programme is based on this theory, RFT will be discussed in detail in the subsequent sections.

**Relational Frame Theory**

Drawing on over four decades of research, RFT advocates suggest that the empirical link between DRR and language arises because they are basically the same phenomenon (i.e., arbitrarily applicable relational responding or relational framing). To explain relational framing, RFT makes a distinction between non-arbitrary and arbitrary forms of relational responding. Most species, from insects to mammals, can relate stimuli based on their formal or physical properties (e.g., picking an object
that is physically larger or smaller than something else, or picking something that is physically the same as another object, as in identity matching). As illustration, a non-arbitrary sameness relational response can be seen if a young child matches a picture (X) of a previously unseen object, in this instance, a trombone, to another picture (Y) of that same object (see Figure 2.1 below).

![Graphical representation of non-arbitrary and arbitrary relational responding.](image)

*Figure 2.1.* Examples of non-arbitrary and arbitrary relational responding (thick lines represent trained relations and dashed lines represent emergent or “derived” relations).

Critically, however, humans display all the hallmarks of a more advanced type of relational behaviour. This enables stimuli to be related, regardless of their physical properties, in ways that were never reinforced in the past, but rather, based on contextual cues that specify the relation. For example, to demonstrate an arbitrary sameness relation, if a child is told that the sound “Trombone” (X) is the name of the object in a picture (Y) and that the written name for “Trombone” is the textual stimulus Trombone (Z), they may derive additional relations (i.e., Y → Z, Z
Theories of Language

➔ Y) which are mutually equivalent without additional training (see Figure 2.1).
That is, X, Y and Z are now mutually substitutable for each other in certain contexts,
such that the selection of any one of the three in the presence of either of the others
will produce reinforcement. According to RFT, this pattern of arbitrary sameness
responding occurs based on the influence of contextual cues such as the phrases “is,”
“called,” or “name of”.

As mentioned previously, sameness is only one type of derived relational
pattern from the RFT perspective. In the case of a difference relation, one person
may say to another “Can you get some fruit in the supermarket, but get something
different than last week, the oranges didn’t taste great”. Based on the inclusion of
the “different” cue, this request would likely evoke a comparison by the listener of
the available fruit options, and the selection of something other than oranges. Social
contingencies might further condition responding; for instance, a return with oranges
might result in reproach and an argument while a return with something other than
this fruit may be approved.

RFT theorists argue that the reinforcement history that has led to this type of
relational responding is what underlies the ability to respond in accordance with
patterns of sameness, distinction, comparison, opposition, analogy, temporality and
deixis. Two characteristics of relational responding or framing that seem particularly
important from an RFT viewpoint are that it is: (i) extremely generative; and (ii) it
can be trained. With specific reference to the first point, several empirical
investigations have deliberately emphasized the generativity of this behaviour (e.g.,
O’Hora, Barnes-Holmes, Roche, & Smeets, 2004; Stewart, Barnes-Holmes & Roche,
2004; Wulfert & Hayes, 1988). Relating to the second point, the ability to derive
relations of various types can be trained, when such responses do not emerge,
following appropriately arranged conditional discrimination training. RFT advocates maintain that framing is learned by neurotypical children via natural everyday environmental contingencies that establish these response patterns (e.g., Luciano et al., 2007). Specifically, parents provide children with multiple exemplars for appropriate responding in accordance with particular stimulus relations.

Consider, for example, the very early history of training responsible for establishing sameness (coordination) relations between a word and a person. In the context of an interaction, a child may be taught to look at a specific stimulus (e.g., a particular person) in the presence of a novel word, such as: “It’s Daddy”, (Child looks at Daddy), (Hear Name X – Look at Person Y). In addition, the child may be taught to produce the best approximation of the name in the presence of the object: (Picture of Daddy shown to the child) “Who is this?” (Child: “Daddy”), (See Person Y – Produce Name X). Initially, children must be explicitly taught each symmetrical relation in both directions (i.e., X-Y; Y-X). However, after receiving sufficient exemplars of bi-directional training, generalization occurs, such that training in only one direction with a novel word-object pair in the object-naming context becomes sufficient for derivation of the relation in the untrained direction (McHugh & Reed, 2008). Effectively, the multiple-exemplar bi-directional training establishes particular contextual cues as discriminative for symmetrical responding. RFT advocates argue that such multiple exemplar training (MET) also facilitates responding in accordance with alternative relational frames (e.g., comparison, opposition). Hence, relational responding is seen as a generalized 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 multiple forms of relational framing, according to RFT, all examples have the following three characteristics: mutual entailment, combinatorial entailment and transformation of stimulus function. These terms are comparable to symmetry, transitivity, and transfer of function in stimulus equivalence, but are broader in that can be employed with relations other than coordination.

Mutual entailment is the most fundamental property of relational framing. It refers to the bi-directional nature of stimulus relations (Hayes et al., 2001). For example, if X is related to Y in a given context, then a relationship between X and Y is entailed. This relationship between the stimuli can be symmetrical (i.e., as in the case of equivalence or coordination), but this may not always be the case. For instance, if we are told that Peter (stimulus X) is younger than Paul (Stimulus Y), it can be derived that Paul (Stimulus Y) is older than Peter (Stimulus X). That is, given a relation in one direction, a second relation is mutually entailed (e.g., if X younger than Y then Y older than X).

Combinatorial entailment occurs when two or more stimulus relations combine to allow derivation of a novel relation. The term combinatorial entailment refers to derived stimulus relations involving two or more sets of relations. Without combinatorial entailment it would not be possible to define the relevant forms of relational frames (Hayes et al., 2001). For example, if, in a given context, X is related to Y and Y is related to Z, then a relation is entailed between X and Z and conversely, Z and X. This property may include, but is not limited to, the transitive relations found in stimulus equivalence. For relations that are mutually entailed, the specified relationship between X and Y always entails a relationship between Y and X at the same level of precision. However, with combinatorial entailment, the
Theories of Language

derived relationship may be less precise than the original relationship. For example, if John (Stimulus X) is different than Maria (Stimulus Y) and Maria (Stimulus Y) is different than Sophie (Stimulus Z), the relationship between John (Stimulus X) and Sophie (Stimulus Z) and Sohie (Stimulus Z) and John (Stimulus X) is unknown. The fact that we do not know the nature of this relationship, in and of itself, is a stimulus relation.

The third characteristic of relational framing is the transformation of stimulus functions. This is said to occur when the functions of one stimulus alter the functions of another stimulus in accordance with the derived relation between the two, without additional training (Barnes, 1994; Dymond & Rehfeldt, 2000; Hayes et al., 2001; Hayes & Wilson, 1993). Transfer of function through equivalence relations is a subtype 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).

For instance, if stimulus X is in an equivalence and/or co-ordination relation with a neutral stimulus Y, and stimulus X acquires fear eliciting functions, these functions may be transferred (via equivalence) to stimulus Y, so that the latter may now elicit fear. For example, if John is frightened of snakes, the word “snake” (which is already in an acquired relation with actual snakes) also evokes fearful sensations. If John hears “un serpent” (the French word for snake) from a friend, but John doesn’t speak French, then he will not experience anxiety. However, if someone tells him that “un serpent” means the same (a contextual cue for a sameness relation) as “snake”, then an equivalence relation between “un serpent” and “snake”
is established. Further, there will likely be transformation of the psychological functions of “un serpent” so that this phrase now elicits anxiety as well. This effect has been demonstrated with many stimulus functions, including conditioned reinforcing functions (Dymond & Barnes, 1995; Roche & Barnes, 1997), discriminative functions (Hayes, Devany, Kohlenberg, Brownstein & Shelby, 1987), elicited conditioned emotional responses (Dougher et., 1994) and self-discrimination functions (Dymond & Barnes, 1994).

To summarize the information presented in this section, from an RFT viewpoint, both DRR and language can be explained as arbitrarily applicable relational responding or relational framing. Relational framing is defined as an overarching operant response class generated by a history of reinforcement across multiple exemplars, and its development underlies the development of language and complex abilities (e.g., problem solving, planning, reasoning etc.), as well as the capacity to show DRR in laboratory experiments. Further, although there are multiple forms of relational framing, all examples have properties of mutual and combinatorial entailment and transformation of function. In the next chapter, studies on DRR that are directly relevant to the application of this phenomenon in the educational arena will be reviewed, specifically teaching generative language to young children with and without developmental delay.

**Summary**

As outlined in this chapter, within the behaviour analytic tradition, several accounts of language acquisition and development have been proposed. The first of these was provided by B.F. Skinner, which, as noted, has had a substantial influence on the basic conceptual framework of behaviour analysis. Nonetheless, at a theoretical and applied level, Skinner’s theory is deficient as an approach to
language (Hayes et al., 2001), most notably because the verbal operants he describes do not facilitate the type of generativity characterising typical language. 

Generativity was then described as being fundamental to the development of fully functional communication. Thus, the development and/or training of this phenomenon was highlighted as being critical. Behaviour analytic research conducted since Skinner has uncovered a phenomenon referred to as DRR which seems extremely relevant as regards understanding generativity in language and has provided the basis for considerable empirical work and theoretical development. Alternative behaviour analytic theories that have attempted to provide an explanation of DRR and the empirical link between DRR and language were described briefly. The most adequate of these and the one on which the current research programme is based is Relational Frame Theory (RFT). 

RFT (Hayes, et al., 2001), explains both DRR and language as AARR or relational framing and research based on this approach has made substantial theoretical and practical advances as regards training the core skills needed for generative language. This concludes the current chapter, which gives a background to RFT as a modern behaviour analytic approach to studying language and cognition. In the next chapter, empirical examples of RFT work on training various types of relational framing in children with and without developmental delay will be described.
Chapter 3

Establishing Derived Relational Responding (DRR) Abilities in Children
Relational Frame Theory (RFT; Hayes, Barnes-Holmes & Roche, 2001) was introduced as a modern contextual behavioural approach to human language in the last chapter. According to RFT, the core process involved in language is a learned, generalized and contextually-controlled type of operant behaviour. This is referred to as derived relational responding (DRR). The RFT conceptualization of DRR as a higher order operant behavior outlines a clear learning pathway for these skills, that is, multiple exemplar training (MET).

Multiple exemplar training (i.e., training a pattern of behaviour in the context of multiple physically dissimilar sets of stimuli) in relational responding skills is: (i) a means by which neurotypical children develop language skills; and (ii) a means of teaching these proficiencies when they are deficient. As alluded to in Chapter 2, there are various contextually controlled relational patterns of responding. Skills within several different frames of DRR have been targeted in both typically developing young children and children with developmental delay. For instance, in a recent literature search, Sánchez and Rodríguez (2014) analyzed the contributions made to date by RFT in children and adolescents with and without developmental disabilities. Following the established criteria, a total of 58 articles were selected between 1986 and 2014, 35 of which included typically developing participants and 23 atypically developing. The relational skills trained in these studies included coordination (Murphy & Barnes-Holmes, 2010; Walsh, Horgan, May, Dymond, & Whelan, 2014), distinction (Newsome, Berens, Ghezzi, Aninao, & W.B. Newsome, 2014), comparison (Barnes-Holmes et al., 2004), opposition (Cassidy, Roche, & Hayes, 2011), and perspective taking (Rehfeldt, Dillen, Ziomek, & Kowalchuk, 2007; Weil, Hayes, & Capurro, 2011). In what follows, empirical examples of the training of these frames will be reviewed.
Training frames of coordination

As aforesaid, relational frames of coordination involve stimuli which are “substitutable” for one another. Moreover, they are also the frames with which most stimulus equivalence research is concerned. Yet, it is difficult to investigate the frame of coordination empirically because it appears to develop quite early in typical language development. For instance, RFT researchers examined the emergence of this frame in a young typically developing child and found that the participant could derive mutually entailed picture-name relations by 17 months, and combinatorially entailed name-sound relations by 24 months (Lipkens, S. Hayes, & L. Hayes, 1993). At the same time, there are several empirical investigations whose work can be considered relevant to the establishment of coordinate framing—at both mutual and combinatorial entailment levels. For instance, in one research study, a toddler (i.e., 15 months) was taught to demonstrate receptive symmetry using MET in bidirectional object-sound/sound-object relations with dissimilar stimuli (e.g., Luciano et al., 2007). First, an object (X) was presented, and was labelled (Y) by the experimenter (i.e., X-Y relations), then (after progressively longer delays) training was provided in the selection of the object from an array (i.e. Y-X relations).

Consequently, the child demonstrated delayed receptive symmetry with novel objects; that is, she could select a specified item from an array (Y-X) for a novel item that had been previously (with a delay) shown to her and labelled by the experimenter (X-Y). After further training in visual-visual conditional discriminations, the participant also showed equivalence.

In an alternative example, Murphy and Barnes-Holmes (2010) examined the development of transfer of function through equivalence relations. The experimenters combined derived relations with manding as a means of facilitating a
more flexible manding repertoire. They used a token board game to contrive conditioned establishing operations for two differently coloured tokens needed to fill the board. First, abstract stimuli X1 and X2 were trained to have discriminative functions for manding the two different colour tokens respectively. Subsequently, participants were trained in X-Y and Y-X conditional discriminations, and then tested for their ability to mand using Z stimuli (i.e., thus showing transfer of the discriminative functions from X1 and X2 to Z1 and Z2 respectively, which in this context was termed “derived manding”). Two participants showed transfer of function immediately. The third, who did not do so, was given MET and after three stimulus sets, the participant showed transfer of function with a fourth novel set.

Finally, in a more recent study, Still, May, Rehfeldt, Whelan & Dymond (2015) used the Relational Completion Procedure (RCP; a task involving dragging and dropping the correct comparison stimulus into a blank box next to the sample stimulus), implemented on touchscreen tablet computers, with 11 children with ASD, aged between 3-12 years. Following formal language and preference assessments, children were first taught to mand via picture exchange for missing items necessary to play with a toy. They then learned (using MET) to conditionally relate the dictated names of the items to the corresponding pictures of the items (X-Y training) and to relate the dictated names to the corresponding printed words (X-Z training). Test probes, in the absence of reinforcement, were presented to determine whether or not participants would mand for the missing items using text exchange (hence demonstrating derived manding/requesting). Probes for spontaneous matching (Y-Z and Z-Y) and labeling (Y-X and Z-X) were also presented in both experiments, based on derived stimulus relations formed among the relations.
Across two experiments, all but one of the participants (P4) demonstrated derived manding and the requisite derived relations.

**Training frames of distinction**

Relational frames of distinction involve responding to the differences among stimuli or events, typically also along a particular dimension. However, in frames of distinction, the relevant dimension is rarely implied. Consider, for instance, the statement “This is not a person of average height”. Based on only this information, it is impossible to ascertain if this person is tall or short. Furthermore, a combinatorially entailed difference relation is unspecified. For example, if you are told that X is different to Y, and Y is different to Z, then X and Z may be the same or different. This type of unspecified relation is a defining property of the frame of distinction.

O’Connor, Barnes-Holmes and Barnes-Holmes (2011) examined mutually entailed arbitrary difference relations in typically developing children and children with autism. In this series of experiments, contextual control was established for emitting a symmetrical (same) response (i.e., trained X1-Y1, tested for Y1-X1), or for emitting an asymmetrical (not same) response (i.e., trained X1-Y1, tested for Y2-X1). The participants were first taught to emit textual responses (Y) to text-based stimuli (X). Following this, they were then trained to select the symmetrically related (same) comparison when presented with the sample and a blue circle, and to select the asymmetrically related (not same) comparison when presented with the sample and a red circle. Once contextual control was established over trained symmetrical (same) and asymmetrical (not same) relations, O’Connor et al. (2011) tested for generalization by establishing arbitrary matching of visual stimuli after training with textual responses. This was followed by testing contextually-controlled
untrained (i.e., derived) relational responses. The results demonstrated that while the typically developing children were able to generalize and demonstrate derived contextually-controlled relational responses to novel stimulus sets, not all the children with autism were able to do so. However, for these children, MET resulted in contextual control over the derived responses. As such, this could be described as an example of MET-based establishment of contextually controlled mutual entailment of “same” and “not same” relational listener discriminations.

An alternative, more recent empirical study investigated the effects of MET on reading comprehension under the contextual cues of “Same” and “Different” (Newsome et al., 2014). Five neurotypical children aged between 9 and 12 years, with strong reading abilities but poor reading comprehension, served as participants. Following training in identifying the categories, features and functions of a variety of common stimuli, participants were next trained in several different relational tasks involving distinction and coordination. For example, participants were asked to identify how a bus and a dog are the same, or how they are different. Participants were also asked questions about activities, such as “How is playing in the park like [or different from] swimming in the pool?” The researchers found that this MET training package improved both the ability to respond to similar relational tasks with novel stimuli, and additionally improved performance on standardized measures of reading comprehension.

**Training frames of comparison**

The comparative frame involves responding to one event in terms of a quantitative or qualitative relation along a specified dimension with another event. Comparative frames may be divided into specific sub-types, such as bigger-smaller, brighter-darker, etc. The different types are, in part, defined by the dimension along
which the relation applies (e.g., size or speed). Comparative frames may also involve quantification of the dimension (e.g., ‘X is more than Y and Y is more than Z’). In comparison relations, the bi-directional relations between stimuli are not symmetrical—for example, mutual entailment would be demonstrated by an individual who, after being taught that X is greater than Y, responds that Y is less than X. In this example, combinatorial entailment might be probed with the addition of a second trained relation such as “Z is less than Y”, for example, and testing for the derivation of “X is greater than Z” and “Z is less than X”.

In an initial study on the training of comparative relations through MET, Y. Barnes-Holmes, D. Barnes-Holmes, Smeets, Strand, & Friman (2004) used abstract stimuli (paper “coins”) and arbitrarily assigned values (being able to buy “more” or “less” sweets with different coins). Three neurotypical children, aged 4 to 6, were taught specific relations among “coins” (e.g., X > Y > Z or X < Y < Z), and were then asked which coin they would or would not bring to the shop to buy sweets. For example, at a three stimuli level, three paper coins were placed in front of the child from left to right (X-Y-Z), and the experimenter asked the following question; “If this coin [pointing to coin X] buys more sweets than this coin [pointing to coin Y] and this coin [pointing to coin Y] buys more sweets than this coin [pointing to coin Z] which would you bring to the shop to buy as many sweets as possible?” Subsequent to extensive training with multiple sets of stimuli (and for one participant, additional training in non-arbitrary comparative relations), all participants were able to demonstrate generalized responding with more than/less than relations. These results have since been replicated and extended (see Berens & Hayes, 2007; Gorham, Y. Barnes-Holmes, D. Barnes-Holmes & Berens, 2009).
In a more recent example, Murphy and Barnes-Holmes (2010) examined the development of derived manding via transformation of functions through a frame of comparison, with both typically developing children and children with ASD. Using a bordgame format, manding was established using stimulus cards (nonsense CVC words/text) for specific amounts (+2, +1, 0, -1, -2) of tokens (smiley faces) while other stimuli (abstract shapes designated X and Y) were established as contextual cues for “more” or “less”. Baseline conditional discriminations were then trained to establish comparative relations (e.g., A>B, B>C, C>D, D>E), using the X/Y stimuli as contextual cues for selecting the appropriate stimulus (the one that is “more” or “less” than the other). Participants were then taught to mand for either +1 or -1 tokens to play the token game, and were subsequently tested for derived mands for +2 or -2 tokens. Two of the seven participants in this study were not able to immediately demonstrate transformation of functions. Thus, the functions were directly trained, by teaching the children to mand with the appropriate stimuli. After MET with two sets of stimuli, these participants were capable of demonstrating derived manding with a novel stimulus set.

Training frames of opposition

The relational frame of opposition requires the abstraction of a dimension along which stimuli can be ordered and distinguished in equal ways from a reference point. In natural language, opposite typically implies the relevant dimension. For example, saying that ‘fast is the opposite of slow’ implies that speed is the dimension along which the related events are to be ordered. Furthermore, RFT suggests that the frame of opposition will emerge later than coordination because the combinatorially entailed relations within frames of opposition are frames of coordination. For
example, if fast is the opposite of slow, and slow is the opposite of quick, then fast and quick are the same (i.e., coordinated), not opposite.

Cassidy and colleagues (2011) used an REP-type protocol to investigate the effects of automated MET patterns of DRR (equivalence, same/opposite and more/less) on intellectual potential, using standard IQ (intelligence quotient) measures as a proxy. These researchers conducted two empirical investigations. In Study 1, four 10–12 year old children were given MET in the relations Same, Opposite and More/Less, across several weekly sessions over a period of 24 months. Baseline and post-intervention IQ scores were assessed using the Wechsler Intelligence Scale for Children – 3rd Edition (WISC-III UK). Matched against four control participants, the training group showed significant improvements in full-scale IQ subsequent to training (i.e., from 105.5 to 132.75 points). In Study 2, a further eight, 11–12 year old children with below average IQs, were exposed to a slightly modified version of the training. Full-scale IQ, measured using the WISC-IV UK, rose by at least one standard deviation for seven of the eight children and this increase was significant. Moreover, an anecdotal four-year follow-up measure recently reported by Roche, Cassidy & Stewart (2013) found that for all seven of the eight participants revisited, IQ rises were maintained and were still significantly higher than the baseline measures taken four years previously.

In an alternative example, Dunne, Foody, Y. Barnes-Holmes, D. Barnes-Holmes, & Murphy (2014) explored an extended sequence of training and testing various relational frames, one of which was opposition in nine children with autism aged four years. Across five phases, participants were trained and and tested with picture based stimuli that targeted alternative dimensions including big/small; long/short; wet/dry; hot/cold; happy/sad; clean/dirty; empty/full; dark/light;
rough/smooth; and heavy/light in non-arbitrary and arbitrary relational responding. For instance, during a non-arbitrary phase, the experimenter may say to the participant “Show me the opposite of the big one” and in the arbitrary phase (where the stimuli were identical), the experimenter might say “If this one is rough, show me the opposite of rough”. Subsequent to MET, opposition responding was established in children who previously lacked this repertoire. Although, this can be seen as demonstrating derived symmetrical responding with antonyms, it is not clear that DRR in a frame of opposition was demonstrated. At the level of mutual entailment, responding in accordance with opposite is symmetrical (if X is opposite Y then Y is opposite X) and is thus indistinguishable from sameness responding. It is not until combinatorial entailment is present (if X is opposite Y and Y is opposite Z, then X and Z are the same) that one could definitively identify this skill as relational responding within the frame of opposition. Moreover, in this study there was no test for the “meaning” of opposite relations, such as a test of transformation of function, or even of non-arbitrary “oppositeness” (e.g., three glasses of water at different temperatures - one cold, one hot and one neutral). If a child cannot pass tests that tap into relevant functions such as these, then it is unlikely that they are responding to “hot” and “cold” as opposite in a meaningful way. As Dunne et al. (2014) did not test for any such functions then, from an RFT perspective at least, their results cannot be seen as a clear demonstration of opposite relations.

**Training deictic frames**

Perspective-taking skills are imperative to an assortment of social interactions (e.g., Baron-Cohen, 2005; Flavell, 2004; Downs & Smith, 2004; Klin, Schultz, & Cohen, 2000). From an RFT viewpoint, the perspective-taking repertoire emerges following a reinforced history of responding relationally to questions such
as, "What will I be doing there", "What would you do if you were me" and alternate questions which require the speaker to change perspective between different references of person (i.e., I versus you), place (i.e., here versus there), and time (i.e., now versus then). McHugh, D. Barnes-Holmes and Y. Barnes-Holmes (2004) developed a protocol to examine these three perspective-taking frames (i.e., I versus you, here versus there, and now versus then) across three levels of complexity (Simple, Reversed, and Double-Reversed). The protocol was administered to a range of participants, ranging from early childhood through adulthood. No consequences were delivered for correct or incorrect responses. The results demonstrated that errors on the task decreased as a function of the participant's age. These findings lend support to the notion that DRR may be the basis of a perspective taking repertoire.

In support of this, Rehfeldt et al. (2007) used MET to train a perspective-taking repertoire in neurotypical participants as well as participants with a diagnosis of ASD, by asking a series of questions. The McHugh et al. (2004) protocol was administered in a conversation format between the experimenter and the participant. For this, the participant had to respond relationally to correctly answer questions such as, "I have a green brick and you have a red brick. If I was you and you were me, which brick would you have? Which brick would I have?". Results of this study suggested that typically-developing individuals via MET, were able to answer these questions and to generalize these responses to novel stimuli. Participants with an ASD diagnosis made significantly more errors than their typically-developing peers when presented with this task. Furthermore, the authors found that DRR repertoire served to facilitate accuracy in this task, also suggesting the importance of a DRR repertoire to the development of perspective taking.
Weil et al., (2011) replicated and extended the aforementioned findings in three younger children aged 4-5 years. Utilising a multiple baseline design across individuals and tasks, deictic relational frames were successfully trained. All participants demonstrated increases in deictic framing that generalized across stimuli, suggesting the acquisition of an operant class. In addition, all of the children showed improvement on Theory of Mind tasks following improvements in deictic performance at the Reversed (e.g., “I have a red ball and you have a blue ball, If I were you and you were me, which ball would I have? Which ball would you have?”) and Double-Reversed (e.g., “I am sitting here on a blue chair and you are sitting there on the black chair. If I was you and you were me and if here was there and there was here, where would I be sitting? Where would you be sitting?”) levels.

Hence, at this point, a growing body of research showing the establishment of DRR skills in children with and without disabilities, across a number of relational frames (e.g., comparison, opposition), has been examined. Most of the empirical examples highlighted above have utilised matching-to-sample (MTS) methodologies (see Cassidy et al., 2011 as an exception). These MTS procedures have been effective in establishing complex patterns of relational responding (e.g. Stewart, Barnes-Holmes, Roche, & Smeets, 2001). However, as will be evident in the next chapter, the dominance of this methodology for studying stimulus relations is not conducive to effective progress for RFT research. An alternative methodology, namely the REP, has been identified as a more efficient way of assessing and training relational framing. This protocol will be reviewed in more detail in Chapter 4.
Summary

The current chapter describes empirical examples of RFT work on training various types of relational framing. These illustrations highlight the prospective of this contextual behavioural approach for educational purposes with both typically developing children and children with a developmental delay.

In the next chapter, the Relational Evaluation Procedure (REP) as an RFT-based protocol will be considered. This methodology has substantial potential as a means of assessing and training relational framing, which, as highlighted, is the principle focus of this research agenda.
Chapter 4

The Relational Evaluation Procedure (REP)
Chapter 3 described studies on establishing Derived Relational Responding (DRR) proficiencies in children. These empirical investigations demonstrated the empirical links between DRR and language, while also illustrating the substantial basic and applied potential of this concept. As detailed, one characteristic of much of the empirical work into DRR within behaviour analysis, broadly, has been the dominance of matching-to-sample (MTS) as a methodology for studying this phenomenon. In a typical MTS procedure, trials commence with the presentation of a sample stimulus. Then, two or more choice stimuli are presented either while the sample remains present or after it has disappeared (referred to as delayed MTS). Selecting the correct choice stimulus contingent on the sample yields access to reinforcement (Doughty & Saunders, 2009), while selecting incorrectly is extinguished or punished. MTS procedures have been employed extensively within behaviour analysis, in both basic research and applied domains, including in the investigation of such phenomena as stimulus equivalence and relational framing.

Consider, for instance, a standard stimulus equivalence procedure using MTS. Participants are initially trained in a series of related conditional discriminations using arbitrary stimuli, such as abstract shapes or nonsense syllables (designated in the write-up of research using alphanumeric code). That is, in the presence of a sample stimulus A1, participants might be trained to choose comparison B1 over the alternative comparison B2. In contrast, they might be trained to choose B2 (not B1) when the sample A2 is present. Hence the relations A1-B1 and A2-B2 are established. Participants might also be trained to select C1 in the presence of B1 (now a comparison rather than a sample) and C2 in the presence of B2. Thus, the additional relations B1-C1 and B2-C2 are established. These would provide baseline conditional discriminations for testing for the emergence of
several potential untrained derived performances. For example, participants might be tested to determine if symmetrical and transitive derived relations emerge. Symmetry is shown when participants respond correctly to tasks that reverse the trained relations (e.g., picking A1 in the presence of B1 [B1-A1]; B1 in the presence of C1 [C1-B1], A2 given B2 [B2-A2]; and B2 given C2 [C2-B2]); and transitivity is shown when participants respond correctly on tasks that require combining two (or more) previously trained relations to produce a novel relation (e.g., picking C1 with A1 or A1 with C1 [A1-C1; C1-A1]); or picking C2 with A2 or A2 with C2 [A2-C2, C2-A2]).

If participants respond in accordance with the above pattern then they have shown stimulus equivalence and more specifically have shown the formation of two, three member equivalence classes or relations (i.e., A1-B1-C1, A2-B2-C2). This is a relatively simple example, however, involving only two comparison stimuli in the MTS protocol and thus leading only to formation of only two classes / relations. The number of classes / relations formed depends on the number of comparisons and research has often involved the use of three, four or sometimes more comparisons, thus resulting in correspondingly larger classes / relations. This is one way in which things can become more complex. Apart from this, additional class members can also be added by increasing the number of conditional discriminations and adding new members that way. For example, in the above example, B stimuli were trained to A stimuli and C stimuli were trained to B stimuli. However, we could have continued to add extra stages in which, for example, D stimuli were trained to C, E stimuli were trained to D etc. Both this and the addition of comparisons would serve to increase the complexity of the final result and with more and / or larger equivalence classes, more complex phenomena can be examined. Nonetheless,
adding to the complexity in this way lengthens the procedures considerably and can put a substantial burden on participants.

Other variations in the MTS for training and testing equivalence classes are the patterns of conditional discrimination training and testing. The above example, in which A is linked with B and then B is linked with C is referred to as a linear procedure. Two other variants are Many-to-One (MTO) and One-to-Many (OTM) procedures. In the MTO procedure, training would involve teaching choice of B comparisons in the presence of an A sample and teaching choice of C comparisons in the presence of an A sample and then equivalence testing would look for relations between B and C stimuli. In the OTM procedure, training would involve teaching choice of A comparisons in the presence of an B sample and teaching choice of A comparisons in the presence of a C sample and once again equivalence testing would look for relations between B and C stimuli. Research has sought to investigate which of these procedures might be more effective in terms of the speed and efficiency of training, as this may be beneficial if more and/or larger classes/relations are to be trained (e.g., Smeets & Barnes-Holmes, 2005; Arntzen & Nikolaisen, 2011; Grisante, Galesi, Sabino, Debert, Arntzen & McIlvane, 2013).

Alternate variables that have been examined as a means of increasing effectiveness are ensuring that conditional discriminations are trained up to high accuracy, fading out reinforcement from 100% to 0% in training before testing, testing simpler derived relations (e.g., symmetry) before more complex ones (e.g., transitivity) and including probes for trained relations amongst probes for tested relations. Procedural variations like these have indeed been shown to have some efficacy in terms of guiding participants to show equivalence responding (e.g., Imam, & Warner, 2014; Minster, Jones, Elliffe, & Muthukumaraswamy, 2006).
Such disparities tend to be most effective given increases in number / size of classes / relations. Despite differences between these variations, however, as number / size of classes / relations increases, the duration of training needed and the encumbrance placed on the participants still increases substantially. Indeed, some of these distinctions, while improving, yield in terms of number of people passing equivalence, obviously only serve to lengthen training still further (e.g., ensuring that people reach 100% accuracy in training or fading out reinforcement). Hence, despite procedural variations, as complexity of the trained / tested performance increases, MTS quickly becomes an ineffective means of training and testing for derived relational responding.

The material just presented details MTS as a means of training and testing for equivalence classes / relations because this has been the key focus of behaviour analysts with respect to the investigation of derived equivalence relations. The MTS has also been drawn upon by RFT researchers, however, as a means of examining DRR more broadly, that is, including not just derived equivalence relations but derived non-equivalence relations also (e.g., different, opposite, more, less etc.). As explained in the previous chapter, RFT explains DRR as generalized contextually controlled relational responding. Hence, RFT sees establishing contextual cues as a key aspect of the investigation of relational responding, not least in order to differentiate between different patterns of DRR. As such, in RFT studies, MTS training typically involves two steps: (1) establishing arbitrary shapes as contextual cues for particular forms of relational responding using non-arbitrary relational training; and (2) using the contextual cues thus established to train participants to relate novel arbitrary stimuli, in particular ways, in the presence of the pre-established contextual cues.
Take, for example, training and testing the derived stimulus relations of Same and Opposite. In the first phase (i.e., non-arbitrary relational training), the purpose would be to establish two arbitrary shapes as having the contextual cue functions of “Same” and “Opposite”. If the shape to be established as Same is presented, then reinforcement will be provided for choosing a comparison that is the same as the sample, whereas if the shape to be established as Opposite is presented, then reinforcement will be provided for choosing a comparison that is as different as possible from the sample along some particular physical dimension (i.e., is opposite from it along that dimension). For example, a contextual cue for Same (i.e., $X$) or Opposite (i.e., $Y$), may be presented on a computer screen, followed by a sample stimulus (e.g., a short line) and then by three comparison stimuli (e.g., of a long line, a medium line, and a short line). Given a short line sample stimulus, in the presence of the Opposite contextual cue (i.e., $Y$), choosing the long line comparison stimulus is reinforced. However, given the Same contextual cue and a short line, choosing the short line comparison is reinforced (e.g., Dymond & Whelan, 2010). Training across multiple exemplars of stimuli that differ along various physical dimensions (shape, color, size etc.) would be conducted, until consistent responding to novel stimuli occurs, in the absence of explicit differential reinforcement.

In the second training phase (i.e., arbitrary relational training), the purpose would be to use the already established contextual cues of Same and Opposite to teach derived stimulus relations. So, while the contextual cues used in the first phase remain the same, the sample and comparison stimuli are now all novel arbitrary stimuli. In the example below, the contextual cue is described first in capitals, followed by the sample stimulus and then by the three comparison stimuli in brackets (correct comparison is in brackets). On a given trial, the notation Same/A1
[B1-B2-N1] would indicate that in the presence of the cue Same (i.e., \( \mathcal{H} \)) and the sample stimulus A1, selecting B1 is reinforced, whereas selecting B2 or N1 is not. Conversely, when the notation Opposite/A1 [B1-B2-N1] is presented, selecting B2 is reinforced, whereas selecting B1 or N1 is not (e.g., Dymond & Whelan, 2010). Trials such as those just represented would effectively train participants that B1 is the same as A1 and that B2 is the opposite of A1. They might, on this basis, derive that B1 is opposite to B2. Thus, if subsequently tested with a trial representable as Opposite/B1 [A1-B2-N2] they will choose B2 as the opposite of B1 rather than either of the other comparisons.

This use of the MTS as a method for establishing contextual cues for various different types of relational responding and then using those cues to train a variety of arbitrary relations, has been used in a multitude of RFT studies interested in investigating different forms of DRR (relational framing) including, for example, “Same”, “Different”, “Opposite” and “More-Than/Less-Than” (e.g., Barnes & Keenan, 1993; Dymond & Barnes, 1995, 1996; Roche & Barnes, 1996, 1997; Roche, Barnes-Holmes, Smeets, Barnes-Holmes, & McGeady, 2000; Steele & Hayes, 1991). Further, highly complex patterns of responding have resulted from this basic procedure (e.g., Steele & Hayes, 1991).

Despite the utility of the MTS procedure, not only with respect to the training and testing of derived equivalence relations, but several other forms of DRR also, several limitations of this protocol have been identified. It was indicated previously, that the MTS protocol can require a substantial amount of training in order to demonstrate DRR, especially as the relational networks involved become larger. In fact this is just one of a number of problematic issues which will now be considered.
RFT researchers (e.g., Lipkens & Hayes, 2009; Hayes & Barnes, 1997; Hayes et al., 2001) have argued for the need to develop new methodologies because the excessive use of MTS may constrain the development of the experimental analysis of DRR. They list a number of reasons for this. First, in MTS, the response of picking, touching or pointing to a comparison in the presence of a sample, explicitly encourages analyses based on the concept of stimulus class (Barnes-Holmes, Hayes, Dymond, & O’Hora, 2001; Hayes & Barnes, 1997). This is problematic for the analysis of derived stimulus relations for a number of reasons including: the concept of stimulus class has narrowed the focus of research to a small subset of relational responses (i.e., Same, Different, Opposite, or More-Than/Less-Than); and the concept of stimulus class precludes a clear understanding of derived stimulus relations at the basic level of process, particularly when examining transformation of stimulus function, one of the fundamental properties of all relational frames (see Barnes-Holmes et al., 2001 for a detailed review). Second, MTS is a methodology best suited for studies of equivalence relations (Hayes & Barnes, 1997) and is constrained by the prior training and testing of a specific set of derived relations. Third, in everyday life, the learning of derived relations rarely formally corresponds to MTS, and research based around MTS procedures may, in some respects, be said to be lacking in ecological validity (Barnes-Holmes et al., 2001). Fourth, as considered previously, some commentators have questioned the extensive training and testing that has been employed in previous MTS research on multiple stimulus relations with adults (e.g., Horne & Lowe, 1997; Stewart, Barnes-Holmes, Roche & Smeets, 2002). This a problematic issue, particularly as relational networks become more complex. As such, the MTS is severely constrained with
respect to the investigation of the potential generativity and complexity of DRR and human language.

Given the potential importance of DRR research for understanding human language and cognition (Dymond & Rehfeldt, 2000; Hayes et al., 2001), it is essential that laboratory procedures are continually refined in order to maximise the efficacy of investigation so that DRR in all its variation, generativity and complexity may be studied as effectively and efficiently as possible (Stewart et al., 2002). This was the purpose of the development of the Relational Evaluation Procedure (REP), a relatively recent alternative to MTS that improves on the older methodology in respects of all of the issues just listed. In what follows, the REP will be introduced as an efficient methodology to assess and train relational responding. First, a history of the REP’s development will be detailed. Following this, a description of the REP itself will be provided, together with a brief review of some key research studies that have used this methodology. After this, a number of research programmes that have evolved from the REP [e.g., Strengthening Mental Abilities with Relational Training (SMART), Implicit Relational Assessment Procedure (IRAP), and the Relational Completion Procedure (RCP)] will be introduced and described. Finally, the potential of the REP as a methodology for assessing and training relational framing in young children will be discussed.

**Development of the Relational Evaluation Procedure**

As part of a broader research program concerned with expanding the range of available methodologies for analyzing language and cognition (see Barnes-Holmes, Healy, & Hayes, 2000; Hayes et al., 2001), an early version of the REP called the Precursor to the Relational Evaluation Procedure (pREP) was first developed in the 1990s. The pREP was based on the Go/No Go Procedure (e.g. D’Amato, &
Colombo, 1985) in which participants were required to respond to one of the choices in a task but withhold a response to the alternative choice. A typical pREP trial involves the presentation of two stimuli on each trial, one conditional stimulus (CS) and one discriminative stimulus (SD). On trials that combined a CS with a positive SD, making a response (for example, pressing the space bar) is reinforced, whereas on trials that combined a CS with a negative SD, it is not. In other words, reinforcement is provided for responding to particular target relations (e.g. A1->B1 and A2->B2), and withheld for responding to non-target relations (e.g. A1->B2 and A2->B1).

An initial experiment using the pREP, (e.g., Cullinan, Barnes and Smeets, 1998) consisted of four conditions, in two of which the baseline tasks were trained using the pREP (Conditions 1 and 2), and in two of which they were trained using a MTS preparation (Conditions 3 and 4). Symmetry (i.e., if A1–B1 then B1–A1) and equivalence (i.e., if A1–B1 and B1–C1, then A1–C1 and C1–A1) were tested with the pREP and with MTS, the pREP before MTS in Conditions 1 and 3, and after MTS in Conditions 2 and 4. The results of this and subsequent studies consistently showed that most participants evidenced pREP symmetry but not pREP equivalence unless: (a) the baseline relations were trained with MTS (Cullinan et al., 1998, 2000); or (b) the pREP task was converted into a MTS task in which the stimulus pairs served as samples with two other “option” stimuli as comparisons (see also, Carpentier, Smeets, & Barnes-Holmes, 2000; Markham & Dougher, 1993; Perez-Gonzalez, 1994), and only when these comparisons were the words “Same” and “Different” (A1 → B1 Same, A1 → B2 → Different) (Cullinan, Barnes-Holmes, & Smeets, 2001). These findings led to the conclusion that the (standard) MTS protocol contains features with pre-experimentally established discriminative
properties for responding in accordance with the relations of “Same” and “Different” (Barnes, 1994; Steele & Hayes, 1991) and that the pREP, lacking these features, leads to the formation of non-separable compounds (Cullinan et al., 2000; Wulfert, Dougher, & Greenway, 1991).

The Relational Evaluation Procedure (REP) was subsequently developed as an extension of the pREP work. The development of this protocol was an attempt to rectify some of the limitations that had been observed when using the pREP. Researchers involved in testing the latter procedure recognised the need to identify possible sources of contextual control that could be incorporated into the procedure (Cullinan et al., 2001). This included training contextual cues for “Same”, “Different” and “Yes” and “No” as an alternative to responding or withholding a response, as required in original pREP trials. Such changes converted the pREP into the REP.

The Relational Evaluation Procedure (REP)

The Relational Evaluation Procedure or REP (see Barnes-Holmes et al., 2001; Stewart, Barnes-Holmes, & Roche, 2004) is a methodology developed by RFT researchers, following dissatisfaction with conventional methods such as MTS and after experimentation with precursor methods (e.g., the pREP), as a means of assessing and training DRR / relational framing under controlled laboratory conditions. The core feature of this protocol is that it induces participants to evaluate the stimulus relation(s) being presented on a given trial, typically using a confirmation or disconfirmation response (e.g., “Yes” and “No”, or analogous terms). In a typical approach, participants may confirm or deny the applicability of particular stimulus relations to other sets of stimulus relations (Stewart et al., 2004). This shifts the focus from stimulus partitioning and picking (with its class
connotations) as evidenced in MTS, to relational specification and evaluation. In early REP work, the functions of cues such as “Same”, “Different” and “Yes” and “No” were established in arbitrary shapes and then used to establish relational networks.

![REP Trials Diagram](image)

*Figure 4.1.* Examples of REP trials. In many (especially early) REP studies, actual English words such as “Same”, “Different”, “Yes” and “No” did not appear; instead, the functions of these words were established in arbitrary shapes using non-arbitrary relational tasks such as those depicted and then those shapes were used to train and test arbitrary relational networks composed of novel arbitrary stimuli (e.g., nonsense syllables). However, the English words (as well as arrows pointing to the correct answers, which also would not appear) are shown here for ease of communication.

Subsequently, two objects in a particular relation with each other (e.g., two same- or differently-shaped objects), a contextual cue (e.g., one of the two arbitrary shapes established as “Same” and “Different”), and the cues for “Yes” and “No” might be presented (see Figure 4.1 for a representation of REP trials). Participants would then be expected to choose the “Yes” comparison when the contextual cue
correctly corresponds to the relationship between the objects (i.e., 'same shape' or 'different shape') while choosing the “No” comparison if it does not.

In order to illustrate standard practice with the originally developed REP more fully, in what follows, an example of pre-training and training, as conducted in a typical study that used this procedure, will be described. This example will utilise a version of the REP employed to study temporal relations (see Barnes-Holmes et al., 2001). It commenced with training on “Before” and “After” relations. On each trial, two arbitrary stimuli were presented, one after the other, in the middle of a computer screen (e.g., LUD → CIG). The presentation of these two stimuli constituted a type of non-arbitrary stimulus event because the two elements were physically related in time (one before or after the other). Following the presentation of this stimulus event, two statements appeared on the screen, one in the lower left-hand corner, and the other in the lower right-hand corner. Both statements contained a stimulus just shown (e.g., LUD), an arbitrary relational contextual cue (e.g., YYY or ZZZ), and the other stimulus just displayed (e.g., CIG). Participants were expected to select one of the two statements, and were then given contingent feedback.

In order to establish YYY as functionally equivalent to the relational cue “Before” choosing the statement LUD (A1) YYY CIG (B1) was reinforced if A1 was presented before B1 at the beginning of the trial. Similarly, to establish ZZZ as functionally equivalent to the relational cue “After”, choosing the statement A1 ZZZ B1 was reinforced if A1 was previously presented after B1. When the “Before” and “After” cues had been trained in this way, they were then tested using new stimuli.

Subsequently, once the functions of the contextual cues were established as just described, participants were trained to respond by either affirming or denying
that the statements “agreed” with one another. If, for example, the statement B1 BEFORE A1 was presented above the second statement A1 BEFORE B1, the two statements clearly did not agree. In this case, the participant received reinforcement for selecting one of two novel nonsense syllables (the syllable thus became functionally equivalent to “No” or “False”). Once the “Yes” and “No” functions had been trained in this way, they could then be tested using new stimuli. When this type of procedure had been established, participants could be trained and assessed on completely novel sets of stimulus relations using the “Yes” and “No” stimuli. Thus, just as with natural language, stimulus relations could be established without the need for explicit, overt responding. Furthermore, any relational stimulus could be taught in this way, if the participant has been exposed to the appropriate initial training exemplars (Barnes-Holmes et al., 2001).

REP Research

Versions of the original REP protocol, similar to that just described, have been used to model and investigate a number of different relations with adult populations including temporal (“Before” and “After”) (O’Hora, Barnes-Holmes, Roche, & Smeets, 2004), distinction (“Same” and “Different”) and analogical frames (Stewart et al., 2004). For instance, O’Hora et al. (2004) employed an REP methodology to examine frames of distinction and temporality in 12 undergraduate college students. Participants were initially exposed to relational training and testing for the aforementioned relations. They were then exposed to a test for instructional control consisting of a number of novel complex probes, that each specified a particular four key response sequence in accordance with derived “Same”, “Different”, “Before” and “After” relations. Subsequent to this, the generativity of this relational responding was examined by exposing participants to a further stage in
which probes employed stimuli from 24 novel sets. The results indicated that eight out of the 12 participants demonstrated responding in accordance with networks of derived “Same”, “Different”, “Before” and “After” relations. Moreover, the control by novel instructions was directly attributable to the initial relational training in which the participants had been exposed.

In an alternative example, Stewart et al. (2004) sought to provide an empirical demonstration of analogy (i.e., analogical relational responding) using the REP. Analogy can be described as the ability to relate two objects and the relations between these objects (e.g. “socks are to feet as gloves are to ___?”). In order to model analogy, the researchers first established abstract shapes as contextual cues for “Same” and “Different” relations and for “True” and “False” responses, respectively, with five adult male participants. After this, these cues were employed to model analogical reasoning and a final test was presented that was designed to examine the relating of combinatorially entailed relations based on non-arbitrary relations (see Figure 4.2). The task required participants to examine nonsense syllables contained in boxes in the center of the screen, and then determine if the relations found therein coordinated with the stimuli found in the boxes presented in the lower section of the screen.

For instance, in Figure 4.2 shown below, LOK and BAB and BAB and a black shape are specified as the same (in the lower boxes). Similarly, URG and WAK, and WAK and a black shape are specified as the same. In this case, therefore, LOK and URG may be defined as participating in a same relation because they are both combinatorially related to black shapes. If similar relational responses occur for BUX and MEY, however, the participant will determine that they are each combinatorially related to differently coloured shapes. Consequently, the relation
between the LOK-URG and BUX-MEY relations is one of difference, not same. On this particular task, therefore, the correct response is “False”, not “True”. Utilising this and alternative similar tasks, the results of this study illustrated that all participants successfully completed the REP model of analogy. Moreover, it demonstrated that an in-principle infinite number of new analogical relations was possible based on this technique. This is because the heart of the REP is the contextual cues. Once these contextual cues are established in the participant’s repertoire, then any stimuli (no matter what their physical form) may be related to any other stimuli in accordance with the cues and a wholly new relational network is created on that basis. This illustrates the utility of the REP for both research and practical purposes.

![Diagram](image)

\textit{Figure 4.2.} A trial from the final stage of Stewart et al. (2004) which involved using an REP methodology to train and test participants in analogical reasoning.

The results of the aforementioned studies (i.e., O’Hora et al., 2004; Stewart et al., 2004) have arguably enabled researchers to explore patterns of relational framing in humans much more effectively than older methodologies, such as the matching-to-sample protocol, could. On the basis of research such as this, the potential of this
protocol as a means of assessing and training relational framing in both research and practical arenas was made apparent. As a result, more recently, the laboratory based REP has inspired the development of further protocols that offer specific advantages for assessing and training relational responding in particular contexts, using both adults and children as participants. These more recent versions of the REP are often more practically oriented than the laboratory REP and thus are not as concerned with modelling language, which requires precisely controlling the history of the establishment of contextual cues used, but instead are concerned with assessing or establishing natural language skills. Such versions would not use arbitrary shapes as contextual cues, but instead natural English language words. However, these protocols feature the same advantages of the REP in other respects and arguably as a result, have been highly effective in their application. One such protocol is the Strengthening Mental Abilities with Relational Training (SMART).

**The REP and the SMART**

The SMART (Strengthening Mental Abilities with Relational Training) protocol is essentially a systematically organised set of REP trials focused on the thoroughgoing assessment and training of core forms of relational framing. In its current instantiation, this online training protocol involves 70 levels that train DRR (specifically “Same/Opposite” and “More/Less”) from relatively simple to relatively complex. Each level includes a training and a test phase. On each trial, participants see one to three statements presented onscreen, in addition to a question which is presented under the statements (see Figure 4.3).
Figure 4.3. Example of a trial on the SMART online training programme.

The response options “Yes” and “No” also appear in the bottom right and left of the computer screen (with their positions counterbalanced across trials). For example, one possible task might present the statement, “HOV is the same as FOV” on the top line, the statement “FOV is the same as DOX” on line 2, the statement “DOX is opposite to DUS” on line 3 and the question “Is DOX the same as FOV?” on the bottom. In this case, as DOX is the same as FOV; thus the correct response would be “Yes”.

Similar to the original laboratory-based REP, SMART involves presenting specific relational terms (e.g., “Similar”, “Opposite”, “More”, “Less”) so that the properties of the relations among the stimuli can be ascertained. In addition, comparable to the laboratory REP, SMART also centres on the establishment of novel stimulus relations. Unlike the laboratory REP, however (and as foreshadowed) the SMART online training programme employs stimuli possessing already established contextual cue functions (e.g., the natural language English words “Yes” and “No”) to assess and train relational framing.
An early prototype of this online method was employed by Cassidy, Roche & Hayes (2011) and investigated the effect of automated multiple exemplar training (MET) of patterns of derived relational responding (“Equivalence”, “Same/Opposite” and “More/Less”) on intellectual potential, using standard IQ measures as a proxy. In Study 1, four typically developing eight to twelve-year-olds were given: (i) stimulus equivalence (SE); and (ii) multiple stimulus relations (MSR) (“Same/Opposite”, “More/Less”) training in a staggered fashion over a period of 24 months. WISC-III UK (i.e., The Wechsler Intelligence Scale for Children-III) IQ measures were recorded at baseline, as well as after both the SE and the MSR training. Matched against no treatment controls, the experimental participants showed significant increases in full-scale IQ following SE training, and a further significant rise following MSR training. In Study 2, a further eight 11- and 12-year-olds, who had been identified by their teachers as having educational difficulties were exposed to an intensive battery of six to fourteen weeks worth of MSR training (administered across approximately nine calendar months). In seven of the eight cases, full-scale IQ as measured by the WISC (IV-UK) rose by at least 1 SD and the improvement was statistically significant at the group level. The data from Cassidy et al. (2011) are highly promising at both a theoretical and a practical level in terms of the apparent effect of training derived relations on intellectual potential. In support of this, Roche, Cassidy & Stewart (2013) used an anecdotal measure four years post intervention and found that for all seven of the eight participants revisited, IQ rises were maintained. Additionally, the results of the initial pilot study have since been replicated (see Cassidy, Roche, Colbert, Stewart & Grey, 2016 and Thirus, Starbrink, Jansson, 2016 as illustration).
Aside from inspiring the SMART, the REP has also led to the development of the Implicit Relational Assessment Protocol (IRAP), which has now become extremely useful and popular in the exploration of participants’ verbal repertoires and is also useful for research into DRR.

The REP and the IRAP

The Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008) is a computer-based procedure that involves the presentation of specific relational terms in a manner that permits the properties of the relations between relevant stimuli to be assessed. For instance, on an IRAP trial, participants might be presented with one of two sample words (e.g., “Mary” or “John”) at the top of the computer screen. Presented directly below this is a target stimulus (e.g., a picture of a doll, or a picture of a car). Participants are required to respond to the relation between the sample and the target by choosing one of two relational options (e.g., “It matches ” or “It doesn’t match”), presented towards the left and right at the bottom of the screen (see Figure 4.4 below). For some trials, progression to the next trial is contingent on selecting a consistent relational response (i.e., in the presence of “Mary” and a picture of a doll, the relational option “It matches” is selected), whereas on other trials, progression is contingent on selecting an inconsistent relational response (i.e., in the presence of “Mary”, a picture of a doll and the relational option “It does not match” is selected). The difference in response latencies across successive presentations of consistent and inconsistent trials provides the critical metric. That is, the average response latencies for participants should be shorter across blocks of consistent trials than across inconsistent trials (see Rabelo, Bortolot, & Souza, 2014). In addition, the extent of
the observed time difference between categorisation is assumed to provide an index of the strength of the verbal or relational responses being assessed.

![Diagram of IRAP trial types](image)

**Figure 4.4.** The four IRAP trial types. Names (John or Mary), toy pictures, and response options are presented all together on each trial.

RFT theorists view the IRAP as measuring fully developed DRR ‘in flight’ under natural language conditions and they are using this procedure to study how the features of DRR change under different conditions. Based on data from IRAP studies conducted over the last decade, they now see relational responding as occurring on a continuum that ranges from “brief and immediate” (referred to using the acronym BIRR) to “extended and elaborated” (referred to as EERR). BIRR is like the earliest reaction, relationally speaking, with respect to a stimulus, based on more deeply ingrained relational learning whereas EERR is not one’s first reaction but one’s more complex relational reaction to one’s earlier reactions; in other words, when a response is elicited by a stimulus, it may be followed by another relational
response, which may occur in response to the stimulus or the response itself. Both BIRR and EERR are forms of DRR, however, which is of course interpreted as a behavioural probability rather than a representation or a mental construct (Hayes, Barnes-Holmes, & Wilson, 2012). Methodologically, the IRAP draws on earlier work of the REP, such that the IRAP was initially called the IREP. After time, the present acronym was adopted because it can be read as “I rap” (as in “I talk quickly”), which conceptually is what the IRAP requires participants to do (Barnes-Holmes et al., 2008).

Similar to the REP, the IRAP involves presenting specific relational terms (e.g., “Similar”, “Opposite”, “More”, “Less”) so that the properties of the relations among the stimuli can be ascertained. However, the IRAP chiefly differs from the REP in that, rather than focusing on the establishment of novel stimulus relations for purposes of laboratory exploration of basic processes, it was designed to assess previously established verbal relations asking participants to respond quickly and accurately in ways that are either consistent or inconsistent with their pre-experimentally established verbal relations with a view to understanding properties of people’s responding in particular psychological domains (e.g., bias, prejudice) or properties of DRR more generally in more complex, naturalistic form.

The basic IRAP effect has since been replicated numerous times, across an array of topics. It has been utilised to detect anti-fat and pro-thin biases (Roddy, Stewart, & Barnes-Holmes, 2010), to measure self-esteem (Remue, De Houwer, Barnes-Holmes, Vanderhasselt, & De Raedt, 2013; Vahey, Barnes-Holmes, BarnesHolmes, & Stewart, 2009) and sexual orientation (Ronspies, Schmidt, Melnikova, Krumova, Zolfagari, & Banse, 2015). It has also been used to measure obsessive-compulsive disorder (Nicholson, McCourt, & Barnes-Holmes, 2013),

As regards the use of the IRAP with children, though research is particularly limited, investigators have examined whether school-aged children display gender-stereotyped attitudes relating to specific toys traditionally associated with either girls or boys (i.e., dolls for girls and cars for boys; see Figure 4.4 Rabelo and colleagues, 2014). They have also examined implicit attitudes to the self versus the opposite gender in typically-developing children and children presenting with ADHD or Dyslexia (e.g., Scanlon, McEnteggart, Y. Barnes-Holmes & D. Barnes-Holmes, 2014). Finally, researchers have adapted the IRAP computer software program into an interactive teaching tool (T-IRAP) (‘T’ for ‘Teaching’) to target relational frames coordination, comparative and opposition in children with Autism Spectrum Disorder (e.g, Kilroe, Murphy, D Barnes-Holmes & Y Barnes-Holmes, 2014). These aforementioned studies suggest the potential benefits of using the IRAP when measuring relational responding in children with and without developmental delay.

One final protocol inspired by the original REP, which is also relevant, is the Relational Completion Procedure (RCP; Dymond, Roche, Forsyth, Whelan & Rhoden 2007, 2008; Dymond & Whelan, 2010).

The REP and the RCP

The RCP has demonstrated efficacy in studying derived relations of “Same” and “Opposite” (Dymond, Ng & Whelan, 2013) and “More” and “Less” (Munnelly, Freegard, & Dymond, 2013). It departs from the REP in a number of procedural
respects. Nevertheless, it is discussed here briefly based on the fact that it is inspired, to some extent, by elements of the REP and thus there is some overlap between the two procedures.

The RCP involves presenting stimuli in sequence from left to right, starting with the sample and followed (1s later) by a contextual cue, a blank space and up to five comparisons. The participant’s task is to drag and drop one of several comparisons into the blank space before confirming each selection (Dymond & Roche, 2013). While this format differs from the standard REP somewhat, key elements, such as the focus on relations between stimuli and on the confirmation of relations are retained. As with a number of other RFT procedures for investigating relational framing at a basic level, the RCP involves a non-arbitrary and an arbitrary relational phase. During the non-arbitrary relational training and testing phase, the functions of contextual cue stimuli are established, and then during the arbitrary relational phase, the cues are used to train and test an arbitrary relational network. For example, in a typical trial in the non-arbitrary phase, the participant might be given a short line as sample stimulus and a short line, a medium line, and a long line as comparison stimuli. In the presence of the contextual cue for “Same”, selecting the short line and emitting the confirmatory response (i.e., evaluating the stimulus relation) is reinforced. On the other hand, given a short line as sample, in the presence of the contextual cue for “Opposite”, selecting the longest line and emitting the confirmatory response is reinforced. During the arbitrary relational training and testing, the response format is identical, except for the use of arbitrary stimuli (Dymond et al., 2013).

Similar to the REP, evaluating the stimulus relation or relations that are presented on a given trial, is a fundamental element of the RCP. Thus, the “Finish
“Trial” and “Start Again” confirmatory responses function in a similar manner to the “Yes” and “No” or “True” and “False” functions of the REP. The RCP differs from the basic laboratory REP in that it uses stimuli possessing already established cue functions (e.g., the English words “Finish Trial” and “Start Again”) to assess or train relational framing. Further there is a drag-and-drop response requirement prior to the confirmatory responses in the RCP, an element not present in typical REP protocols.

The effectiveness of this protocol has been demonstrated in a series of empirical investigations with adults (e.g., Dymond and Whelan, 2010; Dymond et al. 2007, 2008; Dymond & Barnes, 1996; Roche & Barnes, 1997; Steele & Hayes, 1991; Whelan & Barnes-Holmes, 2004) and children (e.g., Walsh, Horgan, May, Dymond, & Whelan, 2014). These findings indicate that the RCP may hold potential as a procedure for studying DRR in both populations.

It is important to note that the children involved in the above mentioned studies, as well as those appearing in the SMART and IRAP studies, were in middle childhood as opposed to earlier childhood; nevertheless, these studies are newly emerging examples of the potential of the REP and its derivatives for work with children.

**REP Research with Children**

The previous sections have highlighted the successful application of REP inspired protocols, namely the SMART, IRAP and RCP in both assessment and training of DRR across a range of studies. As suggested however, in one respect the potential of the REP has not been properly explored. Thus far, this format has mostly involved the assessment and training of DRR in adults (e.g., O’Hora et al., 2004; Stewart et al., 2004; Dymond & Whelan, 2010; Ronspies et al., 2015; Roddy
et al., 2010). REP research with children, both neurotypical (e.g., Rabelo et al., 2014; Cassidy et al., 2011) as well as those with developmental delay (e.g., Walsh et al., 2014; Kilroe et al., 2014) has started to emerge but, there is, as yet, relatively little of it. In addition, the work that has been done has been conducted with older children.

To date, there is no research investigating the potential of the REP with children across various age ranges. Given the importance of assessing and training DRR in children, additional REP research needs to be conducted with this population. As outlined, DRR is a fundamentally important repertoire, underlying language and complex cognitive phenomena. A substantial number of studies have already shown strong correlations between DRR performance and linguistic and cognitive performance and emerging research is also beginning to show the effects of training DRR on intellectual potential (e.g., O’Hora, Pelaez, & Barnes-Holmes, 2005; O’Hora, Pelaez, Barnes-Holmes, Rae, Robinson & Chaudhary, 2008; Cassidy, 2008; O’Toole & Barnes-Holmes, 2009; Cassidy, Roche, & O’Hora, 2010; Gore, Y. Barnes-Holmes, & Murphy, 2010; Cassidy et al., 2011; Moran, Stewart, McElwee, & Ming, 2014; Moran, Walsh, Stewart, McElwee, & Ming, 2015; Tirus, Starbrink, & Jansson, 2016; Parra & Ruiz, 2016; Cassidy et al., 2016). Given the importance of the childhood years for intellectual development, it is of particular scientific and practical importance that we understand and promote the development of DRR during this period. We need to understand the development of particular relational frames as well as their interaction and to facilitate skills in relational framing as a means of overcoming deficits where relevant, as well as of accelerating intellectual abilities. To assess and train DRR, effective instruments are required and the REP has already shown huge potential with respect to adults and older children. As such,
it seems timely to begin to investigate its use in younger children and in children with developmental delay. Effective use of this protocol with these populations might yield substantial rewards with respect to understanding and promoting linguistic and cognitive competencies in individuals for whom such work might be particularly beneficial. Hence the aim of the current research, to expand the use of the REP as a means of assessing and training relational responding in children, including neurotypical children across a variety of ages as well as children with developmental delay. In order to achieve this, a novel protocol, namely the Nonarbitrary Same and Different Relational Evaluation Procedure will be utilised.

**Non-arbitrary Same Different Relational Evaluation Protocol (NSD-REP)**

The NSD-REP protocol was developed for the purpose of this research as a means of assessing and training REP responding in children ranging from two to seven years. It was programmed in Livecode™ and involves three levels. Level 1 evaluates appropriate contextual control over non-arbitrary same and different relations by requiring participants to press a button for either “Same” or “Different” in response to a pair of identical or non-identical pictures. Level 2 assesses for appropriate second order contextual control over non-arbitrary same and different relations. It does this by requiring participants to press a button for either “Yes” or “No” in response to a pair of identical or non-identical pictures presented together with an auditory recording of the words “Same” or “Different”. Level 3 is similar to the latter level but task complexity is increased by including the contextual cue “Not” (see Figure 4.5 as illustration). This particular version of the REP was utilised in all of the empirical investigations presented in this thesis (i.e., Studies 1-9), with the exception of the last study (Study 10), which utilised an alternate version of this protocol (i.e., the SMART).
“NSD-REP Level 1”

“Are these the same or different?”

“NSD-REP Level 2”

“Are these the same (different)?”

“NSD-REP Level 3”

“Are these not the same (different)?”

*Figure 4.5. Illustrations of each of the three levels appearing in the NSD-REP (See Appendix A for additional examples).*
The NSD-REP was designed in advance of the investigations that shape this thesis. Hence, features of this protocol, such as the criteria for advancing from one level to the next and the in-built reinforcement strategies for all three levels were in place prior to the commencement of Studies 1-9. For instance, all NSD-REP levels have the same criterion for number of correct responses needed to pass (i.e., 18/20 = 90% of the raw score, which translates to 8/10 = 80% of the adjusted score). This means that only the participants who attained this pre-defined target progressed to the next level. Conversely, corrective feedback for Level 1 differs from that of Levels 2 and 3. Subsequent to a correct and incorrect response on the first level, the programme presents the auditory stimuli “Yes” or “No” respectively. For Levels 2 and 3 this feedback is replaced by the auditory stimuli “Correct, well done” accompanied by a visual display or “Wrong” with no visual accompaniment as corrective. The rationale for utilising divergent feedback on the first level of the NSD-REP is twofold: (i) to train Level 1 responding; and (ii) to potentially support correct Level 2 and 3 responding when participants are assessed on that stage. In the latter levels, of course, responding “Yes” and “No” to questions about Same and Different is required and as such, Level 1 training that already involves exposure to these cues as feedback to responding might support this repertoire.

The work in this current research programme is preliminary. Some key questions at which it is directed are: (i) approximately when children can begin to be assessed and / or trained on an REP; (ii) how children’s propensity to perform on different REP formats (namely the NSD-REP and the SMART) changes with age; (iii) whether the performance of different age groups of children might correlate with measured linguistic and cognitive potential; and (iv) whether and to what extent
children of different ages, including both typically developing as well as developmentally delayed, can be trained on suitably complex REP formats.

As highlighted in this chapter, a key element of the REP is the ability to respond competently to “Yes” or “No” (or true/false etc.) questions. As such, before considering the use of the REP with very young populations, it would seem important to investigate prior research into yes/no responding in young children. Findings from such work might give some potentially useful insight into typical developmental timing and progression of the emergence of this basic but fundamental skill that could usefully inform subsequent empirical research. An overview of this literature will be provided in the following chapter.

Summary

This chapter described the REP as an efficient way of assessing and training relational framing, both non-arbitrary and arbitrarily applicable. Moreover, it documented how various extensions or evolutions of the REP protocol (i.e., SMART, IRAP and RCP) have proven useful for assessing and training relational responding. Subsequent to an outline of these protocols, one key limitation of REP research thus far was outlined. This limitation is the lack of research with children across a spectrum of age ranges, and especially with young children, the assessment and training of whom it would appear to have considerable potential. The NSD-REP, described above, was developed for this thesis to begin to address this deficit.
Chapter 5

Assessment and training of Yes/No Question responding in children*

* Portions of this chapter have been accepted for publication:
In the previous chapter, the Relational Evaluation Procedure (REP) was introduced as an efficient methodology in assessing and training relational framing. To date, work on the REP has been conducted primarily with adults. It has rarely appeared in the context of studies assessing and training relational responding in children, whether with or without developmental difficulties. Yet, assessing and training this repertoire in children is important for many reasons (e.g., for educational purposes). Given the potential of this methodology, as demonstrated in previous work with adults, more work is needed with children, from early childhood through to later childhood.

One key element required for the REP protocol is appropriate responding coming under the control of the response options “Yes” and “No” (or functionally similar indicators such as “True” and “False”). Apart from its importance in this context, yes/no responding is also a fundamental language skill (Carr, 1982). Hence, it will be informative to consider theory and research on this skill as a background to the exploration of the use of the REP with children. The current chapter details research and theory concerning the development and training of responding to yes/no questions (Y/N-Qs) in young children, both neurotypical and developmental delayed. After reviewing the literature, two theoretical approaches that attempt to explain the emergence of yes/no responding and that have inspired empirical research in this area will be described, namely, Joint Stimulus Control (Lowenkron, 1998) and Relational Frame Theory (Hayes, Barnes-Holmes & Roche, 2001).

The Importance of Y/N-Q Responding

The ability to answer Y/N-Qs is a basic but fundamental language repertoire that plays a pivotal role in social interactions (e.g., Raymond, 2003). The significance of responding to Y/N-Qs is illustrated by the fact that some adaptation
of them has developed in almost every known language system (Sadock and Zwicky 1985).

Across age groups, responding to Y/N-Qs occurs in an extensive array of circumstances to accomplish a widespread multiplicity of pursuits (e.g., Aldridge & Wood, 1998; Race, Ochfeld, Leigh, Argye, & Hillis 2012; Schindler, Kissler, Kühl, Hellweg, & Bengner, 2013; Soderstrom, Ko, & Nevzorova, 2011). One area where they seem to be of particular use, however, is when adults are communicating with children. For example, Y/N-Qs are the principal tactics used to elicit information in medical situations, where doctors are required to obtain information from children in relation to ill-health (e.g., Von Baeyer, Forsyth, Stanford, Watson, & Chambers, 2009), in educational settings, where teachers occasion responses from students (e.g., Edwards & Mercer, 1986) and in eyewitness and forensic settings, where children are interviewed in relation to a crime (e.g., Krahenbuhl & Blades, 2006). For example, Davies, Westcott, and Horan (2000) analysed transcripts of investigative interviews conducted by police officers in the UK. They found that 40% of all the questions posed to four to seven year-olds were closed questions (e.g., yes/no questions). Furthermore, in the developmental literature, researchers regularly use Y/N-Qs to test children’s cognitive and social development (Fritzley & Lee, 2003).

Given the importance of Y/N-Q responding as a fundamental repertoire in human language and in particular the extent to which adults use this repertoire as a means of communicating with children, it is of both theoretical and practical importance to investigate the origins and development of this repertoire in children and to examine processes whereby it can be trained in populations in whom it does not easily emerge.
Y/N-Q in the Cognitive Developmental Literature

Research suggests that by 17 months of age, young children can discriminate Y/N-Qs and will tend to respond to them even though they are unlikely to understand the content of the question. By about 24 months of age they begin to be able to reliably respond to semantic content of Y/N-Qs also as indicated by increasing numbers of correct answers. However even at this age they are still dependent on the presence of concrete lexical items in order to answer correctly (Choi, 1991).

Discrepancies exist in the literature regarding children’s response tendencies toward Y/N-Qs across age, condition, question type and cultural background. One well replicated finding, however, is that children aged between 24 and 36 months tend to show biases of various kinds when answering Y/N-Qs and in particular children closer to 24 than 36 months tend to display a “Yes” bias (e.g., Fritzley, Lindsay, & Lee, 2013; Okanda & Itakura, 2010b; Von Baeyer et al., 2009; Moriguchi, Okanda, & Itakura, 2008).

For example, Fritzley and colleagues (2013) investigated response tendencies of pre-schoolers toward Y/N-Qs about actions. Two hundred children aged 2-5 years were asked questions concerning actions commonly associated with particular objects (e.g., “Did I drink from the cup?”) and actions not commonly associated with particular objects (e.g., “Did I sit on the apple?”). The researchers found that the younger children (i.e., those approximately two years old) displayed a “Yes” bias. This same bias has also been found in previous studies involving Y/N-Qs about objects (e.g., Fritzley & Lee, 2003). In addition, evidence suggests that it is unrelated to culture, interviewer status, or question type. For example, two year olds have been found to show a “Yes” bias to comprehensible yes–no questions.
Yes/No Question Responding

pertaining to both familiar and unfamiliar questions (e.g., Fritzely et al., 2013; Fritzley & Lee, 2003; Okanda, Somogyi, & Itakura, 2012; Okanda & Itakura, 2008) in research conducted throughout the world including North America, Europe, and Asia.

Some work has been conducted to examine possible correlates of “Yes” bias responding in young children. For example, Moriguchi et al. (2008) found that three to four-year-olds’ response bias scores on a yes bias task were significantly negatively correlated with performance on a Dimensional Change Card Sort (DCCS) task and a Picture Vocabulary Test (PVT). On the basis of evidence such as this, Okanda and Itakura (2010b) have proposed that very young children might exhibit a “Yes” bias due to under-development of either verbal or inhibitory control abilities respectively. Further research is needed however.

The results for three-year old children are more varied than for two year olds; however research from a variety of countries including Canada, the U.S., Japan, Vietnam and Hungary has shown that three-year-olds will still exhibit a “Yes” bias under at least some conditions (Okanda, Kanda, Ishiguro, & Itakura, 2013; Fritzely et al., 2013; Fritzley & Lee, 2003; Okanda & Itakura, 2007, 2008, 2010b). In further investigations, Okanda & Itakura (2011) found that three year olds both tended to be incorrect more frequently than six year olds in their answering of Y/N Qs pertaining to both familiar and unfamiliar objects, while also showing significantly shorter response latencies.

Unlike the consistent results from younger pre-schoolers, there were cross-conditional and cross-cultural differences in older preschoolers’ response tendencies to Y/N-Qs. Fritzley and Lee (2003) reported that North American four and five-year-olds did not exhibit a response bias to comprehensible questions but exhibited a
nay-saying bias to incomprehensible questions pertaining to familiar and unfamiliar objects. Japanese and Vietnamese five-year-olds (Okanda & Itakura, 2008) and Hungarian four and five-year-olds (Okanda et al., 2012) have been found to exhibit a nay-saying bias to comprehensible questions pertaining to unfamiliar objects. However, Japanese and Vietnamese four-year-olds exhibited a “Yes” bias when they were asked comprehensible questions pertaining to familiar objects (Okanda & Itakura, 2008), and unfamiliar objects (Okanda & Itakura, 2010b). A “Yes” bias was found in Japanese five- and six-year-olds in relation to familiar objects, but it was significantly weaker than the bias exhibited by three-year-olds (Okanda & Itakura, 2010b). Okanda and Itakura (2010b) and Okanda and Itakura (2011) reported that six-year-olds across cultures tended to exhibit a nay-saying bias to questions pertaining to facial expressions and for unfamiliar object conditions.

**Y/N-Qs in Children with Developmental Delay**

There has been little research into the response to Y/N-Qs among children with developmental delay and all of the work thus far has focused on children with Autism Spectrum Disorder (ASD). Neurotypical children develop the ability to respond to Y/N-Qs through exposure to the natural learning environment, although it is evident from the above, that additional teaching is still required. However, research suggests that children with ASD need additional support in learning to respond to Y/N-Qs questions as well as in generalising these responses to novel contexts. Leaf and McEachin (1999) highlight the importance of training yes/no responding in children with ASD. They argue that ability to answer Y/N-Qs provides a means for communicating preferences, establishes choice making and promotes assertiveness. In addition, teaching communicative rejecting provides the child with a socially acceptable strategy to escape and avoid non-preferred objects.
and activities which may prevent/reduce problem behaviour (Sigafoos, Arthur, & O’Reilly, 2003).

Many of the studies conducted with this group have provided limited information (e.g., specific diagnostic category) concerning the participants involved. Nevertheless, some potentially informative patterns have been identified. For example, the more complex the Y/N-Q and the more unfamiliar the vocabulary or topic, the more likely that echolalia will occur instead of a “Yes” or “No” response (Charlop, 1986; Roberts, 1989; Rydell and Mirenda, 1994). In addition, Oi and Funazaki examined the ability of high functioning ASD children in relation to Y/N-Qs and found: (i) that they could clearly answer simple Y/N-Qs by age seven while lower-functioning ASD children could not; (ii) that they were better able to answer Y/N-Q than “Wh” questions; and (iii) that they had more difficulty answering Y/N-Qs than their typically developing counterparts (Oi, 2010b; Funazaki & Oi, 2013).

Given the necessity of supporting children with ASD in their learning of appropriate responding to Y/N-Qs, a certain amount of research has been conducted towards this end. Some relevant early literature focused on Y/N-Q responding centred on desires and refusals. For example, Hung (1980) described procedures to teach requesting and rejecting to two children with ASD. The participants were taught to produce a vocal verbal “Yes” when offered (‘Do you want?’) preferred food items, and to say “No” when offered non-preferred food items. More recently, similar procedures with adults with severe to profound intellectual disabilities (Duker and Jutten, 1997) were reported. However, in the latter the researchers additionally probed for generalisation across settings and found that the skills did not readily transfer. This is a significant limitation as generalisation is pivotal to the success of any individualised educational programme.
Neef, Walters and Egel (1984) examined the training and generalization of “Yes” and “No” responses from mand to tact conditions. Four children who exhibited “autistic like” behaviour were recorded during tutoring (i.e., tact condition; “Is this a —?”) and embedded instruction (i.e., mand condition; “Do you want a —?”). Results indicated that the participants acquired appropriate Y/N-Q responding during the mand training condition. However, generalized responding to novel stimuli under the tact condition was not observed without the continued use of mand-to-tact stimulus control transfer procedures.

In a more recent investigation, Shillingsburg, Kelley, Roane, Kisamore, and Brown (2009) used an echoic prompt and prompt fade sequence to teach three students with ASD to emit Y/N-Q responding to mands, tacts, and intraverbals. Utilizing a multiple baseline across behaviours, participants demonstrated generalized responding within the same operant class (i.e., to novel or untrained tact stimuli following training with one tact stimulus), but none demonstrated generalized responding across operant classes (i.e., to novel or untrained tact stimuli following training with one mand stimulus). In addition, although generalized Y/N-Q responding to untrained stimuli within the tact condition had been observed, no analysis was provided to account for how this had occurred.

**Theories of Y/N-Q Responding**

Despite the apparent importance of acquiring a repertoire of Y/N-Q answering, it appears that there has been relatively little empirical behaviour analytic work conducted in this domain. Recently, however, this has started to change. In this section two theoretical approaches to explaining generative language will be described. These theories offer expositions regarding the emergence of Y/N-Q answering and have inspired empirical research in this area. One is Joint Stimulus
Joint Stimulus Control

Joint stimulus control (JSC) has been defined as “a discrete event, a change in stimulus control that occurs when a response topography evoked by one stimulus … and preserved by rehearsal, is emitted under the additional (and thus joint) control of a second stimulus” (Lowenkron, 1998, p. 332). For example, imagine that a student is asked to retrieve blue paint. Immediately after the instruction, he echoes it (“blue paint”). As he looks for the item requested, he continues by self-echoing the name of the object. Then when he sees the blue paint, the sight of it evokes saying “blue paint” as a tact, and the topography of this response matches that of the self-echoic. The joint occurrence of these two controlling stimuli then evokes a correct selection response (i.e., picking up the blue paint). Lowenkron (1998) argued that JSC may be used to explain a variety of novel or generative performances, including, for example, derived naming and equivalence. Furthermore, an increasing quantity of research has emerged to support this account in recent years (e.g., Lowenkron, 2006a; Tu, 2006; Esch, Esch, McCart, & Petursdottir, 2010).

Carbone (2015) has used this account to explain the learning of Y/N-Q responding in particular. For example, in one unpublished presentation, he taught a 14 year old male with a diagnosis of PDD-NOS to vocally emit “Yes” or “No” in response to questions about stimulus objects that he could reliably tact. For example, when shown a pencil and asked “Is this a pencil?”, the participant would have to say “Yes.” When shown a pencil and asked “Is this a drum?”, he would have to say “No.” The training stimuli were presented in a randomised order from all stimuli that were not acquired during earlier baseline probes. Once a trained
stimulus was acquired, probes of untrained stimuli were conducted to assess responding to novel stimuli and questions. The results showed a consistent increasing trend in the rate of acquisition of untrained stimuli. A pattern such as this suggests that responding was brought under sources of stimulus control that were common to all trials. Carbone argued that the key process responsible was JSC. More specifically, he argued that saying “Yes” was brought under control of tacting the presence of JSC while saying “No” was brought under control of tacting the absence of JSC (see Figure 5.1 below).

![Conceptual Analysis Diagram]

*Figure 5.1.* Analysis of the Joint Control process with Y/N-Qs.
Thus JSC provides one possible means of explaining the acquisition across exemplars of a generalized Y/N-Q repertoire. It has been suggested that the process of JSC comes to determine correct responding in this and other cases of complex human responding because the controlling relationship between JSC and appropriate responding is incidentally reinforced across exemplars (Lowenkron, 1998). Another recent theory which also explains the acquisition of complex human responding, including Y/N-Q in terms of multiple exemplars, is Relational Frame Theory.

**Relational Frame Theory**

Relational Frame Theory (RFT; e.g., Hayes et al., 2001) explains human language including Y/N-Q as arbitrarily applicable relational responding or relational framing (see Chapter 2 for a full account of this theory, including the nature of relational framing). From this perspective, Y/N-Q is an important repertoire that children learn as they learn to relationally frame and that helps to support their learning of an increasingly complex set of relations and an increasingly extensive relational network. More specifically, the response options “Yes” and “No” can be seen as indicating or referring to relational coherence. Relational coherence or coherent relational responding is defined as relating in a manner that is consistent with what was previously learned within a particular socio-verbal community (Hughes & Barnes-Holmes, 2016). For example, the socio-verbal community with which the current authors and readers are familiar, reinforces relational responding in accordance with the following pattern of comparative relations: If “A is more than B” and “B is more than C” then “A is more than C”. If someone trained within this community is given a particular set of relations of the first two-types (e.g., “a ZID is more than a YIM” and “a YIM is more than a DAX”) and subsequently emits a relational response that does not conform with the taught
pattern (e.g., “a ZID is less than a DAX”), then we might describe this response as incoherent. “Yes” and “No” are response options that a person is taught as a means of referring to or indicating the coherence or otherwise of particular relational networks. For example, imagine that someone has previously learned or derived that “a ZID is more than a DAX”. If they are then asked whether “a ZID is more than a DAX” they can indicate that this coheres with their previously acquired relational network by responding “Yes”; while if they are asked whether “a DAX is more than a ZID”, then they can indicate lack of coherence by responding “No”.

In accordance with this analysis, “Yes” and “No” have been called “relational coherence indicators” (RCI; e.g., Maloney & Barnes-Holmes 2016). Children learn to emit and respond to the RCIs of “Yes” and “No” via multiple exemplar training (MET) provided by the socio-verbal community. The training establishes “Yes” as equivalent to coherent networks and “No” to incoherent networks. This occurs in conjunction with an increasingly complex relational repertoire involving an expanding set of relations and an increasingly extensive relational network. For example, in early childhood the types of relational network in which a child might be asked yes/no questions will often be relatively simple and may involve directly perceptible relations (e.g., “Is the cat on the mat?”, “Is the sky green?”), whereas as they get older, they will ask and answer Y/N-Qs with respect to increasingly complex relational networks (e.g., “Is oxygen a compound?”, “Are you an outgoing person?”, “Does it look like Team X can beat Team Y?”). As such, “Yes” and “No” are used in an increasingly varied set of contexts. Given their extensive use and the variation of the contexts in which they are used, they thus become increasingly generalized.
While “Yes” and “No” are important verbal phenomena, there is little published RFT research as yet into Y/N-Q per se (though see O’Connor, 2004). At the same time, however, Y/N-Q responding has been incorporated into RFT research in an important way, in the context of the Relational Evaluation Procedure (REP), as highlighted in Chapter 4. This is as much as RFT research has done with respect to Y/N-Q so far. However, given the aims of the current thesis, advances in this respect are anticipated.

**Summary**

As highlighted in this chapter, responding “Yes” or “No” represents a key element of the REP. For this reason, research and theory concerning the development and training of responding to Y/N-Qs with children is important and was subsequently detailed. Y/N-Qs occur in a wide variety of contexts to accomplish an extensive array of activities. In addition to being the primary method of eliciting information from children across daily pursuits, Y/N-Qs are also the primary data collection method for young children within the developmental literature. Further, a substantial amount of research has been carried out on response biases to Y/N-Qs. Conversely, relatively little research attention has been given to studying the development of Y/N-Q from a functional analytic perspective. Insights gained from this work would illustrate how best to train young children (whether neurotypical or otherwise) to respond to Y/N-Qs appropriately, so as to facilitate the use of Y/N-Qs in either traditional or more recent (e.g., RFT-based) protocols. Two approaches within behaviour analysis, RFT and JSC, which might be useful for this purpose, were subsequently detailed.

In summary, it is evident from Chapters 4 and 5 that despite the potential of the REP as an assessment and training tool, and its efficacy with respect to
Yes/No Question Responding

predominantly adult populations, research with young children is limited. Moreover, although “Yes” and “No” responding represents an important part of daily interactions for children and is frequently used by adults to communicate with children, there is little research into training this skill, with either neurotypical or developmentally delayed children. The current work will involve exploring the use of the REP with an array of ages of children, including young children as well as older and including both typically developing and developmentally delayed. This work will contribute both to research into the REP as well as research on the development of yes/no responding in children. In the next chapter, research into assessing and training REP performance in very young children will be described.
Chapter 6

Studies 1-5

Assessing and Training Relational Evaluation in Young Children*

In order to adequately characterize a complex phenomenon such as verbal behaviour with precision, scope and depth, innovative methodologies are required to train and test for derived relations (Dymond & Roche & 2013). One such methodology is the Relational Evaluation Procedure (REP). As detailed in Chapter 4, the REP offers several advantages over alternate procedures, one of which is that it can readily accommodate many different varieties of stimulus relations (in contrast to the matching to sample protocol, which is best suited for equivalence studies). Nonetheless, further exploration of the REP as a methodology for use in training relational skills in children has been identified as a deficit in the literature and is thus the aim of the current thesis. In this first empirical chapter, a brief rationale for the current studies is outlined. Following this, the research methods used are described, the results are presented and the findings and their implications are subsequently discussed.

Rationale for Studies 1-5

As aforementioned, the REP or REP inspired protocols (e.g., the Strengthening Mental Abilities using Relational Training or SMART) have been used to establish patterns of derived relations including equivalence (e.g., Cullinan, Barnes, & Smeets, 1998), distinction (e.g., Stewart, Barnes-Holmes, & Roche, 2004), comparison (e.g. Munnelly, Freegard, & Dymond, 2013), opposition (e.g., Cassidy, Roche, & Hayes, 2011), analogy (e.g., Stewart et al., 2004) and temporality (e.g., O’Hora, Barnes-Holmes, Roche, & Smeets, 2004), in both adults and older children.

Given the potential of the REP, efficiently assessing and training relational responding in young children should arguably be a priority. Hence, the work reported in the current chapter, which comprised five studies, sought to: (i) explore
the utility of an REP protocol as an assessment procedure in young (2-4 years old) children, by relating it to standardised measures of intelligence at pre and post-test; and (ii) train appropriate REP-type contextual control over relational responding in this same population. To achieve this, the NSD-REP (Non-arbitrary Same and Different Relational Evaluation Procedure) was employed.

Studies 1 and 5 in the current chapter, investigated the potential of the NSD-REP as a assessment tool by correlating it with measures of linguistic and cognitive performance, specifically, the Preschool Language Scale-4th Edition (PLS4; Zimmerman, Steiner & Pond, 2002) and the Stanford Binet Intelligence Scales-5th Edition (SB5; Roid, 2003). Prior to the inception of this research programme, only one known study had examined the relationship between an REP protocol and intellectual potential in children (i.e., Cassidy et al., 2011). Similar to the Cassidy study, the present series of investigations evaluated performance on the REP, as well as measures of intellectual performance at pre and post-test. Unlike the Cassidy study, however, the cohort of participants utilised in the current study were only assessed on non-arbitrary relations, given that they were much younger (i.e., 2-4 years) than the participants in the cited empirical investigation.

Aside from examining the utility of the NSD-REP as an assessment instrument, Studies 2-4 aimed to explore its use as a potential training protocol in young children. The findings of previous REP-related research with neurotypical children (e.g., Walsh, Horgan, May, Dymond, & Whelan, 2014) suggest that this methodology may indeed be effective for establishing relational responding, both at non-arbitrary and arbitrary levels. Yet, the current work differed from the previous research in a number of important respects. In the abovementioned empirical investigations, the researchers sought to use the REP to assess and train particular
forms of relational responding and thus the children used as participants were older and/or more advanced. Additionally, this research used the REP to target a range of non-arbitrary and arbitrary relations. In contrast, an important aim of the studies described herein, was to investigate the potential of an REP format per se, in very young and less advanced participants. For this reason, the NSD-REP deliberately centred on evaluation of a foundational (i.e., non-arbitrary same and different relations) repertoire, so that if participants were unsuccessful, this failure might be more easily traced to aspects of the format, as opposed to the relational repertoire.

**Study 1**

**Method**

**Participants and setting**

Prior ethical approval for this and all subsequent studies was obtained from the Research Ethics Committee at the National University of Ireland, Galway. Participants were twenty six children (14 female), aged between 2.02 years and 4.11 years \((M = 3.10; SD = 0.77)\) recruited from three separate facilities in County Limerick in Ireland. Two were day care centres in Limerick city that catered for children from birth to five years, while the third was a Montessori school in County Limerick that catered for children aged 3 to 5 years. Participants were included in the study if their parents signed and returned consent forms to the class teacher, they had no special educational needs (as reported by the teacher/parent) and they assented to participation. All participants attending these centres at the time of data collection were included on this basis.

**Materials**

**Yes/No Prescreening Assessment.** This was adapted from Shillingsburg, Kelley, Roane, Kissamore and Brown (2009). It involves 10 picture cards,
illustrating a single stimulus (e.g., ball, car, duck, bear). Each of the ten picture cards is presented in turn and the participant is asked a pre-devised question concerning the identity of the pictured item, to which either a “Yes” or a “No” response is appropriate (five required a “Yes” response and five required a “No” response; see Appendix A). The questions were presented in a randomised order based on a random number generator.

**Stanford Binet 5th Edition-Abbreviated Battery IQ (SB5-ABIQ).** The SB5 (Roid, 2003) is an individually administered test of intelligence and cognitive abilities for individuals between 2 and 85 years of age. The Abbreviated Battery IQ (ABIQ) scale, used in this experiment, consists of two sub-tests: one designated nonverbal (Object Series/Matrices) and one designated verbal (Vocabulary). The “Nonverbal” subtest is designed to assess fluid reasoning skills and includes sequential reasoning and classic matrix items, wherein the person being tested is required to select the best alternative to complete a series or a matrix. The “Verbal” subtest is designed to measure declarative knowledge and in particular vocabulary. Early items require a pointing response (e.g., Item 1; “Touch your mouth”). Mid-level items require single word responses (e.g., Item 12; “What is happening in this picture?” [accompanied by a picture of a person running]). Upper level items require that words be defined clearly (e.g., Item 20; “What is a parrot?”). The SB5-ABIQ takes between 15 and 20 minutes to complete. For the purposes of the current research, total score was based on raw scores; hence the “combined scale” score for the SB5-ABIQ that we report in this study, which we refer to as the SB5 Total (SB5 T) score, is simply the addition of the scores for each of the two sub-scales. The SB5-ABIQ has excellent internal consistency with an average coefficient of .91.
Preschool Language Scales 4th Edition (PLS4). This is a 130-item standardized measure of language skills in children from birth to 6 years 11 months (Zimmermann et al., 2002). It takes 25-40 minutes to administer. The PLS4 is a useful diagnostic and research tool that can be used to identify current comprehension and expressive language skills in addition to measuring changes in language skills over time. The test is individually administered and includes tasks that assess skills in the areas of pre-verbal behaviours, as well as linguistic skills in the areas of semantics, morphology, syntax, integrative language skills, and preliteracy skills. It consists of two sub-scales that assess auditory comprehension (AC) and expressive communication (EC). The AC sub-scale (62 items) 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 sub-scale (68 items) assesses skills including object naming, object description, expression of quantity and the use of grammatical markers. The PLS4 yields norm-referenced test scores (standard scores, percentile ranks, and an age equivalent) for both sub-scales, as well as for total language (TL) score. Test-retest reliability has been reported as ranging between .82 and .95 for the sub-scale (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.

Non-arbitrary Same and Different-Relational Evaluation Procedure (NSD-REP). This was programmed in Livecode™ and presented on a Dell Latitude 2100 touchscreen laptop computer. As aforementioned, it consists of 3 levels. The questions and responses for each of the levels is specified as follows: (1) selection of either the textual stimulus “Same” or “Different” given the auditory prompt “Are these the same or different?” and a pair of either identical or non-
identical pictures; (2) selection of either the textual stimulus “Yes” or “No” given either the auditory prompt “Are these Same?” or “Are these Different?” and a pair of either identical or non-identical pictures; and (3) selection of either the textual stimulus “Yes” or “No” given either the auditory prompt “Are these Not the Same?” or “Are these Not Different?” and a pair of either identical or non-identical pictures (see Figure 6.1). Each level involves 20 trials without feedback, and a response could be scored as either “correct” or “incorrect”. A “correct” response is defined as touching the correct onscreen stimulus while refraining from touching an incorrect stimulus, while an “incorrect” response is defined as touching an incorrect stimulus or touching both correct and incorrect stimuli. The raw score per level was from zero to 20. As a means of controlling for chance responding, the raw score on each level is converted into a final score by subtracting ten and then converting negative scores into zero. This gives a range for a level final score of 0-10. A score of 8 or higher on this index is deemed a pass and allows access to the next level. Level final scores are added to give a total score on the protocol.
Figure 6.1. Examples of the three levels in the NSD REP.

Procedure

**Yes/No Pre-screening Assessment.** All assessments were conducted at a table in a corner of the main activity room used by the day care facility/Montessori. The first session involved administration of the Yes/No assessment, the purpose of which was to ensure that each participant had the capacity to respond to simple yes and no questions. For this, the child sat opposite the experimenter. This assessment was conducted in one sitting and took approximately two to four minutes to complete. No feedback was provided during testing and criterion was set at 90%. All participants attained this criterion and proceeded to the next phase of assessment.
Assessment with PLS4 and SB5-ABIQ. Participants were subsequently assessed on both the PLS4 and SB5-ABIQ. The PLS4 has two separate sub-tests, focused on receptive and expressive communication, respectively. In the case of each child, assessment using this measure was completed in two sessions, each lasting about 16 minutes. In a third session for each child, the SB5-ABIQ was administered, which took approximately 20 minutes to complete.

NSD-REP Assessment. After PLS4 and SB5-ABIQ testing was completed, children were exposed to the NSD-REP protocol. This assessed the evaluation of non-arbitrary same and different relations at three levels of complexity including: (i) same and different; (ii) yes and no; and (iii) yes and no with not.

On each trial of Level 1 (“Same and Different”), two pictures (either identical or non-identical) were presented on the screen accompanied by an auditory prompt (i.e., “Are these the same or different?”). In response, participants were required to press the “Same” button (which always appeared on the lower left of the screen) if the pictures were identical or the “Different” button (which always appeared on the lower right of the screen) if not. No feedback was provided after a response; instead the screen cleared and half a second later the next trial was presented. A total of 20 trials was presented. For Level 2 (“Yes and No”) two pictures (either identical or non-identical) were presented on the screen accompanied by one of two different auditory prompts (i.e., either “Are these the same?” or “Are these different?”). This time participants were required to press either the “Yes” button (which always appeared on the lower left of the screen) or the “No” button (which always appeared on the lower left of the screen). Again, no feedback was provided after a response; instead the screen cleared and half a second later the next trial was presented. A total of 20 trials was presented. Level 3 was identical in all respects to level 2, except
that the auditory prompts were either “Are these not the same?” or “Are these not different?”.

On each of the three levels, responding during a trial was scored as either “correct” or “incorrect”. All levels had the same criterion for number of correct responses needed to pass (i.e., $18/20 = 90\%$ of the raw score, which translates to $8/10 = 80\%$ of the adjusted score). If a participant succeeded in meeting this criterion, they were deemed as passing that level and were then exposed to the next level. If a participant did not attain this target they were deemed to have failed that level. Of the 26 children appraised on Level 1, 16 failed to attain criteria, at which point the assessment ended for these participants. The remaining 10 children were subsequently exposed to Level 2. However, none of these participants were successful in reaching the pre-defined target and therefore, no child was exposed to Level 3 during this study.

In the case of all children, the NSD-REP assessment was completed in one session. The average time to complete the protocol was 4 mins 45 secs (1 minute 55 seconds for Level 1 and 2 minutes 55 seconds for Level 2).

**Results and Discussion**

**Scoring Protocols**

Demographic information and scores for the NSD-REP, PLS4 and SB5-ABIQ assessments are provided in Table 6.1. Scores for the NSD-REP were calculated as outlined in the Materials section. Scores for the PLS4 (raw and age equivalent) and the SB5-ABIQ (verbal and non-verbal), were calculated as per their respective manuals (Zimmermann et al., 2002; Roid, 2003).
Table 6.1.

*Demographic Information, Scores for NSD-REP, PLS4 and SB5-ABIQ assessments*

<table>
<thead>
<tr>
<th>Pt. No.</th>
<th>Age</th>
<th>Sex</th>
<th>Study</th>
<th>NSD-REP</th>
<th>PLS4 T</th>
<th>PLS4 AC</th>
<th>PLS4 EC</th>
<th>PLS4 AE</th>
<th>SB5 T</th>
<th>SB5 NV</th>
<th>SB5 V</th>
<th>SB5 AE</th>
<th>SB5 AB IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.00</td>
<td>F</td>
<td>1,3</td>
<td>8</td>
<td>112</td>
<td>58</td>
<td>54</td>
<td>4.05</td>
<td>30</td>
<td>10</td>
<td>20</td>
<td>4.09</td>
<td>109</td>
</tr>
<tr>
<td>2</td>
<td>3.06</td>
<td>F</td>
<td>1,3</td>
<td>9</td>
<td>113</td>
<td>54</td>
<td>59</td>
<td>4.06</td>
<td>25</td>
<td>8</td>
<td>17</td>
<td>3.11</td>
<td>103</td>
</tr>
<tr>
<td>3</td>
<td>3.00</td>
<td>M</td>
<td>1,2</td>
<td>1</td>
<td>92</td>
<td>44</td>
<td>48</td>
<td>3.03</td>
<td>22</td>
<td>6</td>
<td>16</td>
<td>3.04</td>
<td>103</td>
</tr>
<tr>
<td>4</td>
<td>3.07</td>
<td>M</td>
<td>1,2,3</td>
<td>5</td>
<td>79</td>
<td>40</td>
<td>39</td>
<td>2.07</td>
<td>23</td>
<td>9</td>
<td>14</td>
<td>3.07</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>3.10</td>
<td>F</td>
<td>1</td>
<td>4</td>
<td>103</td>
<td>50</td>
<td>53</td>
<td>3.10</td>
<td>27</td>
<td>9</td>
<td>18</td>
<td>4.03</td>
<td>106</td>
</tr>
<tr>
<td>6</td>
<td>3.06</td>
<td>M</td>
<td>1,2</td>
<td>5</td>
<td>94</td>
<td>47</td>
<td>47</td>
<td>3.04</td>
<td>24</td>
<td>9</td>
<td>15</td>
<td>3.09</td>
<td>103</td>
</tr>
<tr>
<td>7</td>
<td>3.05</td>
<td>M</td>
<td>1,2,3</td>
<td>10</td>
<td>83</td>
<td>40</td>
<td>43</td>
<td>2.09</td>
<td>24</td>
<td>9</td>
<td>15</td>
<td>3.09</td>
<td>106</td>
</tr>
<tr>
<td>8</td>
<td>2.04</td>
<td>M</td>
<td>1</td>
<td>0</td>
<td>65</td>
<td>27</td>
<td>38</td>
<td>2.00</td>
<td>11</td>
<td>3</td>
<td>8</td>
<td>&lt;2.00</td>
<td>88</td>
</tr>
<tr>
<td>9</td>
<td>2.01</td>
<td>F</td>
<td>1</td>
<td>0</td>
<td>68</td>
<td>33</td>
<td>35</td>
<td>2.02</td>
<td>15</td>
<td>4</td>
<td>11</td>
<td>2.04</td>
<td>109</td>
</tr>
<tr>
<td>10</td>
<td>3.00</td>
<td>F</td>
<td>1</td>
<td>0</td>
<td>86</td>
<td>41</td>
<td>45</td>
<td>2.11</td>
<td>23</td>
<td>10</td>
<td>13</td>
<td>3.07</td>
<td>109</td>
</tr>
<tr>
<td>11</td>
<td>2.06</td>
<td>F</td>
<td>1</td>
<td>0</td>
<td>81</td>
<td>40</td>
<td>41</td>
<td>2.00</td>
<td>11</td>
<td>2</td>
<td>9</td>
<td>&lt;2.00</td>
<td>85</td>
</tr>
<tr>
<td>12</td>
<td>4.11</td>
<td>M</td>
<td>1,3</td>
<td>8</td>
<td>109</td>
<td>52</td>
<td>57</td>
<td>4.03</td>
<td>18</td>
<td>8</td>
<td>10</td>
<td>4.03</td>
<td>91</td>
</tr>
</tbody>
</table>
Note: All scores were raw (unstandardized) scores. NSD-REP = Non-arbitrary Same and Different Relational Evaluation Procedure; PLS4 = Pre-School Language Scale 4th Edition; T= Score for total (composite) scale; AC = Auditory Comprehension; EC = Expressive Communication; AE = Age Equivalent; SB5 = Stanford-Binet Intelligence Scales 5th Edition; AB = Abbreviated Scales; NV = Nonverbal; V = Verbal; IQ = Intelligence Quotient.
Correlations Amongst the Measures

Table 6.2 displays a correlation matrix of Spearman rank order correlations among the experimental measures and some of their key sub-scales. This shows a correlation between the NSD-REP and PLS4 total scores ($\rho = .87; p < .001$) as well as between the former and both auditory comprehension ($\rho = .88; p < .001$) and expressive communication ($\rho = .84; p < .001$) PLS4 sub-scales. The NSD-REP also correlated with the SB5-ABIQ Total (SB5 ABIQ T) score ($\rho = .82; p < .001$) as well as with the scores for each of the two sub-scales, $\rho = .79$ ($p < .001$) with the verbal scale and $\rho = .63$ ($p = .0012$) with the non-verbal scale. As such, a key pattern of results in the current study was that the NSD-REP correlated with both the standardised measure of language (i.e., the PLS4) and of cognitive ability (i.e., the SB5-ABIQ).

Table 6.2.

<table>
<thead>
<tr>
<th></th>
<th>NSD-REP</th>
<th>PLS4-REP</th>
<th>PLS4-T</th>
<th>PLS4-AC</th>
<th>PLS4-EC</th>
<th>ABIQ T</th>
<th>SB5-V</th>
<th>SB5-NV</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSD-REP</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS4 Total</td>
<td>.87***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS4-AC</td>
<td>.88***</td>
<td>.98***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS4-EC</td>
<td>.84***</td>
<td>.96***</td>
<td>.92***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB5 Total</td>
<td>.82***</td>
<td>.83***</td>
<td>.87***</td>
<td>.77***</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB5-V</td>
<td>.79***</td>
<td>.80***</td>
<td>.83***</td>
<td>.75 ***</td>
<td>.97***</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB5-NV</td>
<td>.63***</td>
<td>.67***</td>
<td>.72***</td>
<td>.56**</td>
<td>.81***</td>
<td>.70***</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Note: All scores were raw (unstandardized) scores. NSD-REP = Non-arbitrary Same Different Relational Evaluation Procedure; PLS4 = Pre-School Language Scale 4th Edition; T = Score for total (composite) scale; AC = Auditory Comprehension; EC = Expressive Communication; SB5 = Stanford-Binet 5th Edition; ABIQ = Abbreviated Intelligence Quotient; V = Verbal; V = Nonverbal; * = p < 0.05; ** = p < 0.01; *** = p < 0.001.

Table 6.2 also reveals correlations amongst the standardised measures of language and cognitive ability themselves. The SB5-ABIQ Total scale was highly correlated with the PLS4 and its sub-scales; it showed a correlation of .83 (p < .001) with the total scale, of .87 (p < .001) with Auditory Comprehension and .77 (p < .001) with Expressive Communication. The SB5-ABIQ verbal sub-scale was also highly correlated with the PLS4 and its sub-scales. It showed a correlation of .8 (p < .001) with the total scale, of .83 (p < .001) with Auditory Comprehension and of .75 (p < .001) with Expressive Communication. Finally, the SB5-ABIQ non-verbal sub-scale correlated with the PLS4 and its sub-scales. In this case, the correlations were .67 (p < .001) with the PLS4 total scale, .72 (p < .001) with Auditory Comprehension sub-scale and .56 (p < .001) with Expressive Communication.

When examining the relationship amongst the three main measures used, namely, the NSD-REP, the PLS4 and the SB5-ABIQ, the highest correlation was that between the NSD-REP and the PLS4 (ρ = .87); meanwhile the NSD-REP and the PLS4 both correlated at a lower level but still strongly and comparably with the SB5-ABIQ (i.e., ρ = .82 and ρ = .83, respectively). The close correlation of the NSD-REP with both other measures used, reflects the RFT conception that contextually controlled relational responding is the core process involved in language and cognition. The fact that the NSD-REP was correlated slightly more strongly with the dedicated language measure than the measure of cognition may reflect a closer connection between relational responding and focused language tasks.
than between relational responding and a broader range of tasks. Indeed, the pattern of correlations between the NSD-REP and the SB5-ABIQ sub-scales supports this suggestion; the correlation with the verbal sub-scale ($\rho = .79$) was considerably higher than with the nonverbal sub-scale ($\rho = .63$). Interesting, also, are the associations between the NSD-REP and the PLS4 sub-scales. In this case, it can be seen that the NSD-REP correlated more highly with the auditory comprehension sub-scale than with the expressive communication sub-scale and indeed, it correlated more highly with the former than with the PLS4 itself. This might be explained as being a result of the fact that the NSD-REP involves auditory as well as visual stimuli as part of the trials on all three stages and that touchscreen, rather than spoken responses, are required. Of course many of the differences described here are not very large and thus these suggested explanations remain relatively speculative at this time.

**Patterns of responding across age groups**

Also noteworthy are the patterns of responding across differing age groups. As illustrated in Table 6.1, the majority of participants under the age of three and a half failed Level 1 of the NSD-REP without corrective feedback. Conversely, all participants aged four or over passed. Of the participants in between (i.e., 3.05 and 4.00 years), some passed while others failed, suggesting that this repertoire appears to be acquired by typically developing children during this phase. This pattern of responding, perhaps, suggests the start of the acquisition of relatively flexible contextual control (i.e., moving easily between same and different control) during this period. Meanwhile, no participants attained criteria on Level 2 in the absence of feedback. Evidently, the second level involves a more complex level of contextual control, including not just “Same” and “Different” but also “Yes” and “No”. The
latter makes the task even more abstract in that it requires responding to arbitrary relations between the same/different cues and the juxtaposed same/different images and thus this might explain why, in the absence of training, even relatively older children were unable to pass without training.

In summary, the results from Study 1 demonstrate that the NSD-REP correlated highly with measures of language and cognition, which supports the RFT argument that relational responding is a key process underlying language and cognition. The pattern of responding whereby children under 3.05 years generally failed Level 1 while children over 4.00 years all passed suggests the start of the acquisition of relatively flexible contextual control during this period.

**Study 2**

Based on the findings from their initial exposure to the NSD-REP protocol in Study 1, four participants were selected for training in Level 1 responding (this involved selection of either the textual stimulus “Same” or “Different” given the auditory prompt “Are these the same or different?” and a pair of either identical or non-identical pictures) in the context of a controlled intervention study (i.e., a concurrent multiple-baseline design).

**Method**

**Participants and Setting**

Participants 3 (P2.4), 4 (P2.1), 6 (P2.2) and 7 (P2.3) from Study 1 were employed as participants. These four participants (one girl and three boys) were chosen because: (i) they had all failed Level 1 of the NSD-REP; (ii) they were in approximately the same age (i.e., 3.00-3.07 years of age, $M = 3.04, SD = 0.3$); and (iii) they shared a similar school context (i.e., they were all in the same day care
facility in Limerick city). All training and assessments were conducted at a table in the corner of the general education classroom.

**Materials**

The NSD-REP protocol was programmed in Livecode™ and presented on a Dell Latitude 2100 touchscreen laptop computer. Tangible reinforcement (e.g., stickers, erasers, sweets and colouring pages) was also provided at the end of each session for all participants, regardless of phase.

**Target Behaviour and Experimental Design**

A single-subject concurrent multiple baseline design (MBD) across participants (Baer, Wolf, & Risley, 1968) was implemented. Within a MBD across participants, three or more participants are selected and the independent variable, or target intervention, is successively applied to each participant. Experimental control is established when a targeted individuals performance improves while the other participants’ behaviour remains stable. The dependent variable or target behaviour was the percentage of correct responses. Progression of treatment from one behaviour to the next was criterion-based. Participants were required to attain 2 x 90% across 2 consecutive sessions to progress to the next phase of the intervention.

On inspection of the graph however, if there was a decreasing trend of three data points or if data remained stationary across 3 consecutive sessions, an alternative phase was introduced. Each baseline was examined for a stable or downward trend before the next phase was introduced.

**Procedure**

Sessions (involving either assessment or training, depending on the phase in operation for a particular participant) were conducted once a day and consisted of 20 trials. The order in which participants received their sessions on a particular day was
randomized across all four. Prior to a session, the instructor asked the participant concerned if they would like to play the same and different game on the computer. If the child agreed, he or she was brought individually to the teachers desk in the corner of the room.

**Phase I-Baseline.** During this phase, participants were exposed to Level 1 of the NSD-REP as described in Study 1. This involved the presentation of one (randomized) stimulus set. There was no feedback given for correct or incorrect responses during baseline sessions.

**Phase II-Training.** After three baseline sessions, Participant 1 commenced training. This phase differed from baseline in that auditory feedback (i.e., “Yes” for correct and “No” for incorrect) was provided after each response. After a correct response, the computer presented the auditory stimulus “Yes” after which, the screen cleared for 0.5 seconds before the start of the next trial. For incorrect responses, the computer presented the auditory stimulus “No” and the program automatically stopped to allow the researcher to provide feedback. She did this by repeating the auditory (question) prompt and providing a further vocal and gestural prompt to highlight the correct answer. During this phase, a stimulus set was used for two consecutive sessions before a novel stimulus set was introduced. This switching, which was independent of accuracy on a particular stimulus set, was effected to try to bring responding more under the influence of the non-arbitrary same and different relations involved than features of particular stimulus sets. The first participant (P2.1) commenced training after three baseline sessions. Each subsequent participants intervention started after the previous participant had achieved criterion and a minimum of three sessions after the previous participant’s intervention had commenced.
Phase III-Baseline. When participants reached criteria in the training phase (i.e., 90% correct across two consecutive sessions), they were re-tested with the stimulus set from Phase I. This occurred one week post-intervention. Identical to the initial baseline, no feedback was provided during this phase.

Phase IV-Follow-up. During this phase, generalised responding to non-arbitrary same and different relations was assessed by introducing a novel stimulus set and providing no feedback during testing. This was carried out two weeks subsequent to the end of training.

Results and Discussion

Figure 6.2 displays the effects of training on the percentage of correct responses across four phases (i.e., baseline, training, baseline and follow-up) for four participants. Training produced an increase in the percentage of correct responses for all participants. P2.1 and P2.3 demonstrated criterion level responding within two sessions of intervention and ultimately only required three training sessions to meet criterion fully. They maintained these high scores when the baseline stimulus set was re-administered one week later and sustained them also in a test with a novel stimulus set administered two weeks after training had finished. P2.2 and P2.4 completed training in four and five sessions respectively. Furthermore, scores for these participants were comparable to those of P2.1 and P2.3 with respect to baseline and follow-up (generalisation) performance.
Figure 6.2. Study 2. Percentage of correct responses for Level 1 of the NSD-REP during PI baseline (BL), PII intervention (INT), PIII baseline (BL) and PIV follow-up (FU), across each participant. Intersection lines indicate progression to the next phase of the study. Numbers in brackets after participant numbers indicate number of that participant in Study 1.
The results of this second study demonstrate the efficacy of the NSD-REP format for training simple REP type relational responding in young children. Specifically, this study successfully established correct responding to Level 1 of the protocol, which required responding to “Same” and “Different” as textual stimuli in response to identical and non-identical pictorial stimuli, respectively. Given the success with Level 1 of the protocol, the aim of the next study was to examine whether a similar approach might be used to establish correct responding on the second level of the NSD-REP.

**Study 3**

To extend the results of the previous investigation, Study 3 aimed to investigate the possibility of teaching participants to successfully show NSD-REP Level 2 responding, which involved answering “Yes” or “No” to non-arbitrary same and different relations. Once again a MBD was employed.

NSD-REP Level 2 differs from Level 1 in a number of important respects. The primary difference is the level of contextual control involved. In Level 2, participants were required to come under the control not just of “Same” and “Different” but also “Yes” and “No”. Specifically, they were required to press “Yes” or “No” given either the auditory prompt “Are these same?” or “Are these different?” and a pair of identical or non-identical pictures. This is a particularly important level as responding “Yes” or “No” in the context of coherent or non-coherent relations is a core part of the typical REP methodology. As such it was important to investigate whether this might be readily trained.

Corrective feedback for Level 2 is also different from that for Level 1. Subsequent to a correct response, the programme presents the auditory stimulus “Correct, well done” accompanied by a visual display consisting of a grid of
coloured stimuli including stars, small animals and shapes. For incorrect responses, the programme presents the auditory stimulus “Wrong” with no visual accompaniment. This replaces the auditory stimuli “Yes” and “No” as corrective feedback from Level 1. One key reason for the difference is that for Level 1 it was originally decided to use feedback that would not only train Level 1 but would also potentially support correct Level 2 responding when participants were assessed on that stage. In the latter level of course, responding “Yes” and “No” to questions about Same and Different is required and as such, Level 1 training that has already involved exposure to these cues as feedback to responding might support this repertoire.

**Method**

**Participants and Setting**

Participants 1 (P3.4), 2 (P3.2), 4 (P3.5), 7 (P3.7), 12 (P3.3), 14 (P3.1), 15 (P3.6) and 21 (P3.8) from Study 1 were employed as participants in the present study, (P4 [P3.5] and P7 [P3.7] were also in Study 2 as P2.1 and P2.3, respectively). These eight children, who included four girls and four boys, ranged in age from 37-59 months ($M = 49.5, SD = 8.0$). They came from two separate facilities in Limerick city, including a city based day care facility and a Montessori school (for a description of both, please see Study 1).

**Materials**

As in Study 2, the NSD-REP protocol was administered using a touch screen Dell Latitude 2100 computer. Tangible reinforcers (e.g., stickers, erasers, sweets and colouring pages) were also provided at the end of each session for all participants, regardless of phase.
Target Behaviour and Experimental Design

The current study was implemented across two educational settings. A nonconcurrent MBD across participants was used. This meant that each baseline length was pre-determined (i.e., 3, 6, 9, 12 sessions). While there are drawbacks to utilising this experimental design, including that the inference one can make regarding the independent variable is weakened (Harris & Jenson, 1985); and that it has reduced control of threats to internal validity (Hayes, 1985), the non-concurrent MBD has the advantage of greater flexibility in recruitment of participants and testing location. Given the latter, nonconcurrent multiple baseline designs are often recommended for research in educational settings (Harvey, May, & Kennedy, 2004), as was the case in the current study.

The target behaviour was correct NSD-REP Level 2 responding (i.e., choosing either “Yes” or “No” as appropriate in response to trials involving the juxtaposition of non-arbitrary same or different and an auditory cue of “same” or “different”). The same pass/progression criteria as used in Study 2 were employed here also.

Procedure

A maximum of one session (of 20 trials) was conducted per day, per participant (involving either assessment or training, depending on the phase in operation for a particular participant). Apart from this constraint, sessions were conducted as frequently as possible across the two facilities involved. Prior to a session, the instructor asked the participant concerned if they would like to play the same and different game on the computer. If the participant agreed, he or she was assessed and / or training individually in a quiet location where the session could be conducted undisturbed. Identical to Study 2, participants were provided with
tangible reinforcement at the end of each session (i.e., FR20), regardless of phase. Social reinforcement was also provided during the training phase (i.e., Phase II).

**Phase I-Baseline.** During this phase, participants were exposed to Level 2 of the NSD-REP, as described in Study 1, with one (randomized) stimulus set (i.e., involving twenty pairs of visual stimuli). There was no feedback given for correct or incorrect responses during baseline sessions.

**Phase II-Training.** After three baseline sessions, P3.1 and P3.2 commenced training. This phase differed from baseline in that (auditory and visual) feedback was provided for correct and incorrect responding. After a correct response, the computer presented the auditory stimulus “Correct, well done” accompanied by a visual display, consisting of a grid of coloured stimuli including stars, small animals and shapes. After this, the screen cleared for 0.5 seconds before the start of the next trial. For incorrect responses, the computer presented the auditory stimulus “Wrong” with no visual accompaniment and the program automatically stopped to allow the researcher to provide feedback. She did this by repeating the auditory (question) prompt and providing a further vocal and gestural prompt to highlight the correct answer. During the training phase, a stimulus set was used for two consecutive sessions before a novel stimulus set was introduced. P3.3 and P3.4 started training after 6 baseline sessions; P3.5 and P3.6 started training after nine baseline sessions; and P3.7 and P3.8 began training after twelve baseline sessions.

**Phase III-Baseline.** When participants reached criterion in the training phase (i.e., 90% correct across two consecutive sessions), they were re-tested with the stimulus set from Phase I. This occurred one week post-intervention. As in the (Phase I) baseline, no feedback was provided during this phase.
Phase IV-Follow-Up. This final phase assessed for generalization and maintenance of NSD-REP Level 2 responding by assessing performance on this level with a totally novel stimulus set in the absence of any feedback. This assessment was carried out two weeks after the end of training.

Results and Discussion

P3.1 took a total of three sessions to reach criterion, with a mean of 1m 32s per session; P3.2, P3.3 and P3.8 took a total of five training sessions, with a mean of 1m 56s, 1m 38s and 1m 30s per session per session respectively; P3.4 and P3.6 took a total of four sessions, with a mean of 1m 27s and 1m 24s per session; P3.5 took a total of 10 sessions with a mean of 1m 32s per session; and P3.7 took a total of 7 sessions, with a mean of 2m 01s per session.

Figure 6.3 shows the results of the nonconcurrent MBD procedure employed in the present study to show the effect of NSD-REP Level 2 (Yes and No) training. Similar to Study 2, there were four phases (i.e., P I baseline, P II training, P III baseline and P IV follow-up). As can be seen, all participants maintained a relatively stable performance during Phase I (Baseline). Phase II (Intervention) resulted in an increase in the percentage of correct responses in the case of all but one of the participants. P3.1, P3.2, P3.4, P3.6, P3.7 and P3.8 demonstrated an immediate increase in the percentage of correct responding and quickly reached criterion. Furthermore, they maintained criterion level responding in Phase III (Return to Baseline) and Phase IV (Follow Up).

P3.3 met criterion in Phase II within three sessions but failed to maintain criterion responding in Phase III (Return to Baseline). Hence, Phase II was repeated. After reaching criterion within two sessions, Participant 3 was then re exposed to
Phase III (Return to Baseline) and passed and subsequently also passed in Phase IV (Follow Up).

P3.5 was exposed to Phase II after nine exposures to Phase I; however this participant failed to reach criterion after ten sessions of training and thus, in accordance with pre-defined criteria, at this point participation was ended. It should be noted that this participant had previously taken part in Study 2, which trained Level 1 responding. This participants responding on Level 2 seemed to be influenced by the training received in the previous study, whereby, the participant vocalised “Same” and “Different” when responding to the “Yes” and “No” button stimuli.

The main purpose of Studies 2 and 3 was to investigate the efficacy of multiple exemplar training (MET) for improving the performance of young children on various levels of the NSD-REP as a methodology for training relational responding. Overall, the results of these studies support the conclusion that young children aged between three and four can be taught to respond sucessfully on this protocol. Given the success with this age group, one question that arises is whether an even younger cohort (in the 2.5 – 3 year range) with a potentially less developed repertoire (as measured using similar standardised testing procedures) might also be trained. This was the focus of Study 4.
Study 3 Data

\[ \text{Same/ Diff with Yes and No (Test); } \heartsuit \text{ Same/ Diff with Yes and No (Train)} \]
Figure 6.3 Study 3. Percentage of correct responses for Level 2 of the NSD-REP during phase I baseline (BL), phase II intervention (INT), phase III baseline (BL) and phase IV follow-up (FU), for each of the eight participants involved. Intersection lines indicate progression to the next phase of the study. Numbers in brackets after participant numbers indicate number of that participant in Study 1.

Study 4

Study 4 aimed to teach NSD-REP Level 1 responding (i.e., non-arbitrary same and different relations) to two young children in the 2.5 – 3 year range. This study differed from Studies 2 and 3 in that non-automated as well as automated instruction was employed. When it was found during initial testing that participants could not respond on the NSD-REP, a multi-phase table top procedure was implemented prior to re-introducing the computer protocol. Non-automated or table-top procedures offer several advantages when working with very young children. These include the interactive nature of the tasks, the use of social reinforcers, the flexibility of the procedure and sessions of varying duration (Dymond, Rehfeldt, & Schenk, 2005). Moreover, non-automated procedures have been used successfully to assess and train relational responding in children with and without disabilities (e.g., Y Barnes-Holmes, D Barnes-Holmes, & Smeets, 2004; Berens & Hayes, 2007; Dunne, Foody, Y Barnes-Holmes, D Barnes-Holmes, & Murphy, 2014).

Method

Participants and Setting

Participants 23 (P4.2) and 25 (P4.1) from Study 1, both of whom had failed Level 1 of the NSD-REP, participated in this study. P4.1 (female) was 2.11 years while P4.2 (female) was 2.08 years ($M = 2.95; SD = 0.02$). Both participants were recruited from a daycare facility in Limerick city (see Study 1). As may be seen in Table 6.1, both P4.1 and P4.2 had lower scores than any of the participants in Study
2 (which had also focused on training Level 1) on a number of key measures including PLS4, PLS4 age equivalent, SB5-ABIQ and SB5-ABIQ verbal.

**Materials**

As in Study 2, the NSD-REP protocol was administered using a touch screen Dell Latitude 2100 computer. In addition, a number of stimulus sets, each made up of ten picture pairs (including pictures of animals, shapes, common objects and cartoon characters), were used during table-top based stages of assessment and training. Tangible reinforcers (e.g., stickers, erasers, sweets and colouring pages) were also distributed as per Studies 2 and 3.

**Target Behaviour and Experimental Design**

The target behaviour was similar to Study 2; however, the core intervention used a table-top rather than a touch-screen, computer-based methodology. In its essence, the design could be considered an ABA design conducted for each of two participants, where the A phase was performance on a computer-presented task while the B phase involved remediation using a table-top format. At the same time, however, both the A and B phases involved sub-phases of testing and training and thus in that sense, the design is more complex than a simple ABA design. The same pass/progression criteria for assessment and training as outlined for previous studies were employed here also.

**Procedure**

**Phase A - NSD-REP Level 1 Computer-Based Assessment & Training.**

The first section of this phase involved assessing participants on Level 1 of the NSD-REP protocol without feedback. If participants failed criterion, they were then moved onto NSD-REP Level 1 Training. The latter was the same as the previous part except that it included feedback. If participants passed training (i.e., $2 \times 90\%$
accuracy across 2 consecutive sessions), they would then be moved into the previous (assessment) part. If they failed training (i.e., there was a decreasing trend of three data points or if data stayed stationary across 3 consecutive sessions), participants’ were exposed to table-top assessment and training.

**Phase B – Table-Top Assessment & Training.** This phase, which involved table-top assessment and training, included a number of sub-phases.

**Table-top Sub-Phase 1.** During this sub-phase, participants were exposed to a quasi-random sequence of table-top-presented same and different trials. During each trial, two pictures (either identical or non-identical) were laid face up on the table in front of the participant and the researcher then presented an auditory $S^D$ (i.e., “Are these the same or different?”). A correct response involved saying “Same” if the pictures were identical and “Different” if not. No feedback was provided after a response; instead the pictures were collected and those for the next trial presented. A total of twenty trials (ten “same” and ten “different”) was presented. If participants passed then they would be re-exposed to the NSD-REP Level 1 Training. If not, then they were moved on to Phase 2 of the table-top protocol.

**Table-top Sub-Phase 2.** This sub-phase was essentially the same as Sub-phase 1 except that participants received feedback (either “Correct” or “Wrong”). It involved two sub-sub-phases: (i) where trials were organised in alternating three-trial sequences involving either all “Same” or all “Different” trials; and (ii) where trials were presented quasi-randomly in pairs involving just a single “Same” and a single “Different” trial. If participants passed sub-phase 2 then they would be given Sub-phase 1 again. If they failed then they would be moved to Sub-phase 3.

**Table-top Sub-Phase 3.** This sub-phase involved using an all visual conditional discrimination format with a sample picture which was given to the
participant and two comparison pictures, one identical to the sample and one non-identical, laid out in front of the participant. On each trial, the researcher asked the participant to either “find a picture that is the same” as the sample or “find a picture that is different” from the sample. This sub-phase involved three sub-sub-phases; (i) same trials only; (ii) different trials only; and (iii) a quasi-random mix of same and different trials.

After completing Phase B, participants were re-exposed to Phase A.

Results and Discussion

Figure 6.4 displays the data for each of the two participants in Study 4, with the upper and lower panels showing the data for P4.1 and P4.2 respectively. P4.1 took a total of 42 sessions with a mean of 3m 23s per session; P4.2 took a total of 76 sessions with a mean of 3m 33s per session. P4.1 failed both testing and training in the A (computer-based) Phase and thus was exposed to the B (table-top) Phase. This participant appeared to progress relatively quickly through the table-top procedures but did not show immediate improvement on the Phase A probes. After re-exposure to additional Phase B training however, this participant showed improvement in Phase A probes and then passed both training and testing of Phase A. P4.2 failed both testing and training in the A (computer-based) Phase and thus was exposed to the B (table-top) Phase. P4.2 required greater levels of exposure to B sub-phases than the first participant. Eventually, however this participant met training criteria for these phases, after which she went on to successfully complete both training and testing in Phase A. Both participants also maintained their scores when a generalisation stimulus set (without feedback) was administered one week later.
Study 4 Data

- Phase A (i) NSD-REP (no feedback)
- Phase A (ii) NSD-REP (with feedback)
- Phase B (i) TT quasi randomised S & D trials (no feedback)
- Phase B (ii a) TT three trial S or D trials (with feedback)
- Phase B (ii b) TT quasi randomised S & D trials (no feedback)
- Phase B (iii a) TT, MTS S trials (with feedback)
- Phase B (iii b) TT, MTS D trials (with feedback)
- Phase B (iii c) TT, MTS S & D trials (with feedback)
Study 4 Data

- Phase A (i) NSD-REP (no feedback)
- Phase A (ii) NSD-REP (with feedback)
- Phase B (i) TT quasi randomised S & D trials (no feedback)
- Phase B (ii a) TT three trial S or D trials (with feedback)
- Phase B (ii b) TT quasi randomised S & D trials (no feedback)
- Phase B (iii a) TT, MTS S trials (with feedback)
- Phase B (iii b) TT, MTS D trials (with feedback)
- Phase B (iii c) TT, MTS S & D trials (with feedback)

Phase A

Phase B
Figure 6.4 Study 4. Upper panel: Percentage of correct responses for level 1 of the NSD-REP during various phases (see legend) for participant 4.1. Lower panel: Percentage of correct responses for level 1 of the NSD-REP during various phases (see legend) for participant 4.2.

It should be noted that participants in this study required a significantly longer time period to acquire the required repertoire (i.e., responding to non-arbitrary same and different relations under the contextual control of “Same” and “Different” cues) that participants in previous studies (i.e., Studies 2 and 3). In Study 2, which also assessed Level 1 NSD-REP responding, training lasted between three and five sessions for participants who were on average 3.05 years (range, 3.00 - 3.07) with mean age equivalent scores of 3.00 years (range, 2.07 - 3.04) on the PLS4 and 3:07 years (range, 3.04 – 3.09) on the SB5. Conversely, in Study 4, training lasted 42 and 76 sessions respectively for participants aged 2.10 years (2.08 and 2.11 respectively) with mean age equivalent scores of 2.04 years (2.02 and 2.06) on PLS4 and 2:08 years (2.05 and 2.11) on the SB5. This represents a difference of five months in actual age, as well as eight and 11 months respectively in mean age equivalent PLS4 in SB5 scores. Further, it suggests relatively less developed repertoires on the part of the participants in Study 4, which might account for the difference in training times. It should be noted that for these participants, reaching the pre-defined training target may have been the result of acquisition of these repertoires through exposure to extra-experimental training in addition to training effects. Hence, while these participants did attain criterion responding, participants with comparable repertoires might need additional supports in place or more intensive training if they are to acquire the needed repertoires relatively quickly and based on training alone rather
than extra-experimental influence. Further exploration of REP responding with very young children is necessary prior to affirming this assertion.

In addition, unlike the aforementioned studies, this investigation also involved the use of non-automated (i.e., table-top) procedures as well as automated procedures (i.e., computer-based instruction). Using both protocols was warranted given that participants were unable to respond to the NSD-REP until table-top instruction had been provided. Though there are advantages associated with table-top procedures for young children (e.g., social reinforcement, flexibility), one of the limitations in the current study was the lack of interobserver agreement during the non-automated sessions (something that is not required using computer-based protocols) to rule out the possibility of experimenter cuing. Ultimately, for the training of certain types of (e.g., basic relational) repertoires, it is desirable (e.g., for purposes of experimental control, efficiency, etc.) that children might be able to learn through automated procedures and hence the aim in this study to facilitate that through an initial exposure to analogous table top training.

Though the methodology in Study 4 differed from Studies 2 and 3 to accommodate a younger cohort of participants, overall these investigations highlight the utility of the NSD-REP as a training protocol in very young children. The next study in this chapter examines pre to post-test alterations in the standardised linguistic and cognitive assessments.

Study 5

Study 5 involved an analysis of the data from Studies 2-4 as a means of providing a preliminary test of the effects of NSD-REP training on the intellectual potential of the participants involved in those studies.
Method

Participants

This study involved analysis of the data of twelve participants (numbered as per Study 1, that is, P1, P2, P3, P4, P6, P7, P12, P14, P15, P21, P23, P25) across Studies 2-4.

Materials

The SB5-ABIQ, the PLS4 and the NSD-REP were used for this Study (for a detailed description, please refer to Study 1, pp. 93-95)

Procedure

All participants were re-administered the SB5-ABIQ, the PLS4 and the NSD-REP subsequent to the end of training. For participants 4 and 7, who received training on two levels (i.e., 1 and 2) across two studies (i.e., 2 and 3), the measures were not examined until both children completed Study 3.

Re-assessment with PLS4 and SB5-ABIQ.

The receptive and expressive assessment of the PLS4 were completed in two sessions that lasted 22 and 23 minutes respectively. This was followed by the re-administration of the SB5-ABIQ, which took approximately 21 minutes to complete. No feedback was provided for correct or incorrect responses during these sessions. However, participants received tangible reinforcement at the end of all sessions.

NSD-REP Assessment. After PLS4 and SB5-ABIQ re-testing were completed, children were re-exposed to the NSD-REP protocol. Identical to Study 1, all levels had the same criterion for number of correct responses needed to pass (i.e., 18/20 = 90% of the raw score, which translates to 8/10 = 80% of the adjusted score). If a participant succeeded in meeting this criterion, then they were deemed as passing that level and were then exposed to the next level. If a participant did not attain this
target they were deemed to have failed that level. Of the 12 children appraised on Level 1 in this study, all met criteria to and were subsequently exposed to Level 2. Eight participants were successful in reaching the pre-defined target on Level 2 and were then assessed on Level 3. However, none of the participants met criteria for this Level. All three levels of the NSD-REP assessment were completed in one session. The average time to complete the protocol was 6 mins 28 secs (58 seconds for Level 1 and 1 minute 31 seconds for Level 2 and 3 minutes 49 seconds for Level 3).

***Results and Discussion***

Table 6.3 shows outcome data based on the training of the twelve participants from Studies 2-4. It should be noted, in particular, the portion of the table labelled “Difference” which shows recorded differences in participant ages, NSD-REP scores and “Age Relative Performance” scores for standardised tests from pre-test to post-test. It is to be expected that all twelve participants would show improvement on the NSD-REP, because they were trained to criterion on one or more stages. In addition, however, they also tended to show improvement relative to their ages on the other measures. For the PLS4, all participants, except 1, showed improvement (median = 7) while for the SB5, eight out of twelve did so (median = 1.5). A Wilcoxon’s signed ranks test comparing pre and post age related performance scores for both PLS4 and SB5 was conducted. The result for the PLS4 was significant \( (z = -2.93, p < .01) \) while the that for the SB5 was not \( (z = -1.68, p = \text{NS}) \). Spearman’s rank tests were also conducted looking for correlations in the improvement amongst the three measures (i.e., NSD-REP, PLS4 and SB5) but no significant correlations were found.
We also investigated possible differences in ARP improvement on both PLS4 and SB5 measures from pre- to post-test for (i) those trained on NSD-REP Levels 1 versus Level 2 and (ii) younger (i.e., < 3:08 years) versus older (i.e. >3.08 years) participants. Tests showed (i) a significant difference (p < .05) between those trained on Levels 1 and 2 of the NSD-REP and (ii) between the younger and older cohorts of participants in relation to PLS4 performance, with greater increases associated with Level 2 training and the older children that participated respectively. In the meantime, there were no significant differences recorded with regard to ARP on the SB5.
### Table 6.3.

**Demographics and scores at pre and post-test for SB5-ABIQ, PLS4 and NSD-REP**

<table>
<thead>
<tr>
<th>No. In Study</th>
<th>Age in Months</th>
<th>SB5 AE in Months</th>
<th>SB5 ARP in Months</th>
<th>PLS4 AE in Months</th>
<th>PLS4 ARP in Months</th>
<th>NSD-REP Score</th>
<th>Age in Months</th>
<th>SB5 AE in Months</th>
<th>SB5 ARP in Months</th>
<th>PLS4 AE in Months</th>
<th>PLS4 ARP in Months</th>
<th>NSD-REP Score</th>
<th>SB5</th>
<th>PLS4</th>
<th>NSD-REP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48</td>
<td>57</td>
<td>9</td>
<td>53</td>
<td>5</td>
<td>8</td>
<td>56</td>
<td>62</td>
<td>6</td>
<td>65</td>
<td>9</td>
<td>20</td>
<td>-3</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>42</td>
<td>47</td>
<td>5</td>
<td>54</td>
<td>12</td>
<td>9</td>
<td>51</td>
<td>57</td>
<td>6</td>
<td>70</td>
<td>19</td>
<td>21</td>
<td>1</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>40</td>
<td>4</td>
<td>39</td>
<td>3</td>
<td>1</td>
<td>44</td>
<td>59</td>
<td>15</td>
<td>47</td>
<td>3</td>
<td>19</td>
<td>11</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
<td>43</td>
<td>0</td>
<td>31</td>
<td>-12</td>
<td>5</td>
<td>51</td>
<td>51</td>
<td>0</td>
<td>51</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>42</td>
<td>45</td>
<td>3</td>
<td>40</td>
<td>-2</td>
<td>5</td>
<td>51</td>
<td>51</td>
<td>0</td>
<td>52</td>
<td>1</td>
<td>10</td>
<td>-3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>41</td>
<td>45</td>
<td>4</td>
<td>33</td>
<td>-8</td>
<td>1</td>
<td>50</td>
<td>51</td>
<td>1</td>
<td>48</td>
<td>-2</td>
<td>10</td>
<td>-3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>59</td>
<td>51</td>
<td>-8</td>
<td>51</td>
<td>-8</td>
<td>8</td>
<td>63</td>
<td>59</td>
<td>-4</td>
<td>68</td>
<td>5</td>
<td>21</td>
<td>4</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>
**REP Training of Young Children**

<p>| | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>56</td>
<td>61</td>
<td>5</td>
<td>51</td>
<td>-5</td>
<td>10</td>
<td>59</td>
<td>73</td>
<td>14</td>
<td>77</td>
<td>18</td>
<td>20</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>15</td>
<td>53</td>
<td>53</td>
<td>0</td>
<td>52</td>
<td>-1</td>
<td>9</td>
<td>56</td>
<td>66</td>
<td>10</td>
<td>63</td>
<td>7</td>
<td>20</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>21</td>
<td>57</td>
<td>71</td>
<td>14</td>
<td>53</td>
<td>-4</td>
<td>10</td>
<td>60</td>
<td>73</td>
<td>13</td>
<td>75</td>
<td>15</td>
<td>21</td>
<td>-1</td>
<td>24</td>
</tr>
<tr>
<td>23</td>
<td>33</td>
<td>29</td>
<td>-4</td>
<td>26</td>
<td>-7</td>
<td>0</td>
<td>37</td>
<td>39</td>
<td>2</td>
<td>35</td>
<td>-2</td>
<td>9</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>25</td>
<td>35</td>
<td>35</td>
<td>0</td>
<td>30</td>
<td>-5</td>
<td>0</td>
<td>38</td>
<td>40</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

Note: Ages (in months), scores on various measures at pre-test and post-test and differences between pre-test and post-test scores; SB5 = Stanford Binet 5th Edition-Abbreviated Intelligence Quotient; PLS4 = Pre-School Language Scale 4th Edition; NSD-REP = Non-arbitrary Same Difference Relational Evaluation Procedure; “Mths” = “Months”; AE = “Age Equivalent”, the score expected to be attained on a particular test for a particular age of child based on normative data; ARP = “Age Relative Performance”, which is the AE score in months minus age in months.
General Discussion

Several previous studies have demonstrated the potential of the REP protocol as a means of assessing and training relational responding. Most of this work has been conducted with adults or at least school age children (e.g., Stewart et al., 2004; Cassidy et al., 2011). The current series of studies aimed to extend this work by investigating the utility of the REP as a means of assessing and training relational responding in very young (i.e., pre-school) children. In Studies 1-5, typically developing pre-school children ranging from 2 to 4 years of age were assessed and trained on the NSD-REP, whose core focus was contextual control of same and different cues over non-arbitrary same and different relations. In Study 1, a correlational study involving a sample size of 26, NSD-REP performance was shown to correlate highly with performance on mainstream measures of language and cognition (i.e., the SB5 and PLS4). In Studies 2-4, cohorts of children who did not meet criterion on a particular level of the NSD-REP, were successfully trained to pass that level. Finally, in Study 5, an analysis of the data from Studies 2-4 demonstrated that, relative to their ages, participants showed improvement on measures of cognitive and linguistic ability.

When examining the correlations amongst the NSD-REP, the PLS4 and the SB5-ABIQ in Study 1, the highest correlation was found between the NSD-REP and the PLS4; and that both these measures correlated at a slightly lower level with the SB5-ABIQ. The close correlation of the NSD-REP with both assessments reflects the RFT conception that contextually controlled relational responding is the core process involved in language and cognition. As aforementioned, the higher correlation with the measure of language, may reflect a closer connection between relational responding and focused language tasks than between relational responding
and a broader range of tasks. It was argued that the pattern of correlations between
the NSD-REP and the SB5-ABIQ sub-scales supports this suggestion; the correlation
with the verbal sub-scale was higher than with the nonverbal sub-scale. The
correlations between the NSD-REP and the PLS4 sub-scales were also noted. In this
case, it was seen that the NSD-REP correlated more highly with the auditory
comprehension sub-scale than with the expressive communication sub-scale and
indeed it correlated more highly with the former than with the PLS4 itself. As
suggested, this might be a result of the inclusion of auditory as well as visual stimuli
as part of the trials on all three stages of the NSD-REP and the fact that touchscreen
rather than spoken responses were required. While these suggestions remain as yet
speculative, the broad pattern of correlations supports the RFT argument that
relational responding is a key process underlying language and cognition.

The finding that relational responding correlates well with performance on
measures of language and cognition is not new. In fact, this is now a relatively well-
established empirical finding based on work with both adults and children and
involving individuals with developmental delay as well as typically developing
populations (e.g., O’Toole & Barnes-Holmes, 2009; Moran, Stewart, McElwee &
Ming, 2014; see Stewart and Roche, 2013, for a review of previous research).
Importantly, it should be noted that non-arbitrary same and difference relations
themselves were not the target of assessment/training in these studies; rather, the
target was appropriate REP-type contextual control over such relations. Non-
arbitrary same and difference relations were chosen as the focus of the NSD-REP
because this is a relatively early performance marker and thus participant’s failure to
perform on one or more levels would be less likely to be based on failure in this skill
than on failure to come under appropriate REP-type contextual control. The REP-
type contextual control referred to can be characterised as follows: Level 1: arbitrary contextual control (S/D) over non-arbitrary relations; Level 2: Second order (Y/N + S/D) arbitrary contextual control over non-arbitrary relations; and Level 3: Negation controlled second order (Y/N + S/D) arbitrary contextual control over non-arbitrary relations. Patterns of contextual control such as this are crucial for REP performance and thus these were the focus of the protocol used. This might explain why participants failed on at least some of the levels of the NSD-REP, even though those involved had language and cognitive equivalence scores that might suggest that they should pass simple tests of non-arbitrary relations (especially fundamental relations such as same and different). Once it is seen that the latter protocol is not a test of simple non-arbitrary same and different, but of more complex skills, then this pattern is less surprising.

One feature of the current work which may be unique is the relative ease of administration and thus practical utility of the protocol used to gauge level of relational responding. The correlations with mainstream measures were high, suggesting that the protocol is indeed a useful measure of relational responding as a core aspect of intellectual performance. At the same time, the assessment was brief (e.g., average assessment time of less than six minutes) and unproblematic (all children could be tested) as a means of measuring relational performance, even in very young children. As such, this protocol or something similar may be a useful gauge in future studies, behavioural or otherwise, not just of relational responding but of intellectual performance more generally.

In addition to constituting a successful means of assessing relational responding, this study also illustrates the utility of the NSD-REP as a means of training this repertoire. As reported, over the course of a number of experiments, the
protocol was employed successfully to train patterns of relational responding in very young children, including a number of children under the age of three. This suggests that the REP is a potentially useful method for training relational responding in very young populations; furthermore, it recommends, in particular, the NSD-REP as a useful exemplar of the REP.

Apart from showing improvement in relational responding based on REP training, this study also extended previous work by indicating that such training could bring accompanying improvement on linguistic and cognitive tests. Specifically, participants who received training (and showed improvement) on the NSD-REP also tended to show improvement relative to their ages on a number of other standardised measures (see Study 5). All participants, except 1 showed improvement on the PLS4 while eight out of twelve did so on the SB5. In addition, a comparison of pre and post age related performance showed a significant improvement for the PLS4 though not for the SB5. This adds to previous data (e.g., Cassidy et al., 2011) suggesting the effect of training generalised relational responding on intellectual performance. In this particular case, however, the participants, ranging in age from 2 to 4 years, were much younger than in previous studies, in which children were, at youngest, eight years of age. The current protocol was adjusted to focus on non-arbitrary relational responding as opposed to derived relational responding and this probably constitutes a useful adjustment for relational training in very young children. Further work involving relational training is recommended to further utilise and investigate non-arbitrary relations. It should also be stated that the suggested improvement in the current study was calculated based on age relative performance as gauged via normative standards used in standardised tests as opposed to improvement in performance based on comparison with a control
group and as such it should be considered as a tentative indicator only. Further, better controlled studies will be needed to substantiate these findings.

Taken together, however, the results from Studies 1-5 support the efficacy of the NSD-REP as means of assessing and training relational responding in very young children.

**Summary**

In sum, the current series of studies, the first undertaken in the context of the present programme of doctoral research, provides insight into the prospective utility of the REP (and more specifically, the current version of this protocol, the NSD-REP, or something similar) with very young children. Indeed, the data suggest overall, this methodology is a useful tool for assessing and training contextual control over non-arbitrary simple relations with this age group. These findings thus highlight the potential of the REP in future investigations of non-arbitrary and potentially arbitrary relations in very young children. Collectively, the results add to a growing literature that supports the RFT perspective on the importance of relational responding vîs-a-vîs linguistic and cognitive performance. Further, they support the use of the REP as a means of training and testing relational responding.
Chapter 7

Study 6

Assessing and Training Relational Evaluation in Children with Autism

Spectrum Disorder*

* This research was completed in co-operation with a student completing a Masters of Science (MSc) in Applied Behavior Analysis, whose involvement contributed to the completion of a research thesis in partial fulfilment of the requirements of her course.


134
The previous chapter described the first five empirical investigations completed as part of this doctoral thesis. This research explored REP performance in an age range (2-4 years; i.e., early childhood) at which children were acquiring the necessary repertoire to respond on an REP. A multi-level protocol focused specifically on contextually controlled non-arbitrary Same/Different relations, namely the NSD-REP (Non-Arbitrary Same and Different Relational Evaluation Procedure), was utilised to evaluate REP potential in these studies.

Study 1 assessed and scored participants’ performance on this protocol, as well as standardised tests of cognitive ability and language. This allowed categorization of children by repertoire so as to facilitate identification of individuals who might be trained in subsequent studies. It also enabled correlation of NSD-REP performance with intellectual potential more broadly. As predicted, findings showed strong correlations between NSD-REP performance and the alternate measures. Studies 2-4, which followed, targeted particular repertoires for training based on the outcomes of the first study. Using multiple baseline designs (MBDs) with different cohorts, results demonstrated that NSD-REP type relational repertoires, ranging from relatively simple to relatively complex, could be effectively trained in young children. Finally, Study 5 involved an analysis of the data from Studies 2-4 as a means of providing a preliminary test of the effects of NSD-REP training on the intellectual potential of the participants involved in those studies. The findings indicated that, relative to their ages, participants showed improvement on measures of cognitive and linguistic ability.

Given the relative efficacy of the NSD-REP with young typically developing children, one possible domain for extension of this work is to children with developmental delay. This is the purpose of the research presented in the current
REP training of Children with Autism

Chapter; specifically, this work investigated the use of the NSD-REP as an assessment and training methodology for children with Autism Spectrum Disorder (ASD). To commence, the rationale for Study 6 will be provided and the methodology will be detailed. After introducing and reporting the results, the possible implications are then discussed.

**Rationale for Study 6**

A core diagnostic feature of ASD is the presence of qualitative impairments in communication (American Psychiatric Association, 2000). As such, teaching language is one of the critical components of effective intervention programmes for children with this neurobiological disorder, and behaviour analytic approaches place a special emphasis on the development of such skills (e.g., see Sundberg & Michael, 2001). Researchers and practitioners have typically employed the verbal operants described by Skinner in his book ‘*Verbal Behavior*’ (1957) to develop procedures for targeting language skills (e.g., Greer & Ross, 2008; Sundberg & Michael, 2001). This has proven very useful in developing functional communicative repertoires in children with ASD (e.g., Cautilli, 2007; Prelock, Paul, & Allen, 2011) and has become an integral part of early behavioural intervention (e.g., Sundberg & Partington, 1998).

Despite its utility, teaching language utilising the Skinnerian approach is time and labour intensive. Moreover, this approach has been unsuccessful in furnishing an effective understanding of the skills that underlie the phenomenon of generative language, that is, the ability to produce or understand novel sentences (Hayes, Barnes Holmes & Roche, 2001). As highlighted in Chapter 2, Relational Frame Theory (RFT), a modern behaviour analytic account of language and cognition, has convincingly accounted for language development in terms of the learned operant of
derived relational responding (DRR; e.g., Barnes-Holmes, McHugh, & Barnes Holmes, 2004; Hayes et al., 2001; Hayes, Fox, Gifford, Wilson, Barnes-Holmes, & Healy, 2001). Two characteristics of DRR 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. Hence, language training programmes for individuals with ASD that focus on establishing derived stimulus relations would be beneficial (Rosales & Rehfeldt, 2007). Research suggests that a sizeable percentage of children with ASD do not easily learn this key form of responding via typical socio-verbal interactions and hence their language level remains below that seen in their peers (Moran, Stewart, McElwee, & Ming, 2010). However, with targeted training, children with such deficits might acquire this repertoire (Moran et al., 2010).

Extensive research has examined the emergence of derived relations in individuals with ASD (see McLay, Sutherland, Church, & Tyler-Merrick, 2013). In the majority of these studies, these skills emerged following the teaching of specific relational responses (Murphy & Barnes-Holmes, 2010); however, for some, antecedent conditions needed to be modified in order for these responses to emerge (McLay et al., 2013). Research from the literature has also demonstrated that an individual’s level of verbal competence correlates positively with the ability to derive arbitrary relations (e.g., Barnes, McCullagh, & Keenan, 1990; O’Connor, Rafferty, Barnes-Holmes, & Barnes-Holmes, 2009; O’Hora, Pelaez, & Barnes-Holmes, 2005). Further, a growing body of research suggests that interventions that teach DRR improve verbal ability (see Barnes & Rehfeldt, 2013 for an overview).

In Chapter 4, the REP was introduced as an RFT methodology that researchers have developed explicitly for training DRR. This protocol enables
researchers to explore patterns of relational framing in humans much more effectively than in older methodologies. Up until now, the REP protocol has mostly involved the assessment and training of derived relational responding in adults (e.g., Stewart, Barnes-Holmes, & Roche, 2004). Given the advantages of the REP and its potential as an educational methodology, it seems useful to investigate its use in children with ASD and to examine how to train children who show deficits in this area.

One example of a study that has used an REP inspired protocol to increase relational responding skills in individuals with ASD was conducted by Kilroe, Murphy, D Barnes-Holmes & Y Barnes-Holmes (2014). These researchers adapted the Implicit Relational Assessment Procedure (IRAP), a derivative of the REP, into an interactive teaching tool (T-IRAP) (‘T’ for ‘Teaching’) to target relational frames with four children with ASD aged eight to ten years. A modified MBD was used to compare participants’ relational learning in terms of speed and accuracy during table-top teaching and T-IRAP teaching. The table-top procedure commenced with all participants simultaneously, and the T-IRAP was introduced at stepwise time intervals across the four participants. The targeted relations were coordination, comparative and opposition and frames of coordination, comparative and opposition. Results revealed that the T-IRAP was successfully adapted to teach all targeted relations. Additionally, general greater speed and accuracy in relational responding was shown for all four participants during T-IRAP teaching compared with table-top teaching.

The current study differed from Kilroe et al. (2014) in two ways. First, the focus of the research presented in this chapter was to investigate the capacity of the children involved to acquire REP-type responding itself and in particular the ability
to evaluate relational coherence by responding either “Yes” or “No”. As such, the current study employed participants with a less advanced repertoire than those in Kilroe et al. (2014). Second, the current study also sought to explore the utility of the NSD-REP as an assessment procedure in the context of children with ASD by relating it to a standardised measure of language, specifically, the Preschool Language Scale-4 (PLS4; Zimmerman, Steiner & Pond, 2002); no similar correlation between REP performance and intellectual potential as measured on a mainstream measure was made by Kilroe et al. (2014). It was expected that, similar to the findings of Studies 1-5 in the current thesis, there would be a positive correlation between REP score and language ability.

**Method**

**Participants and Setting**

For the assessment phase, nine children (8 male, 1 female), aged between 7 and 11 years ($M = 8.09; SD = 1.05$) took part. These were full time attendees at a local school for individuals with ASD. They each had an independent diagnosis of autism provided by a psychiatrist or a licensed clinical psychologist.

All children initially assessed had normal vision. All had vocal verbal repertoires (Skinner, 1957) including textual responding (i.e., reading via word recognition). All were assessed for a history of seizures with the intention to exclude positives in case of photosensitive reaction to the computer screen; none were excluded on this basis. None had previously taken part in similar research. Children were included based both on parental consent and the verbal assent of the children themselves. Participants were informed that they could withdraw from the experiment at any time. Assessment and teaching was conducted on an individual
basis, and the main researcher was present with each child throughout all procedures. If participants showed signs of distress, procedures were terminated.

The language and relational skills assessments were carried out during regular teaching hours, in an unoccupied classroom. The researcher and participant were seated opposite each other at a small table. The relational skills training intervention was carried out in the general education classroom at a small desk where the researcher and participant sat beside each other. Other students in the classroom were engaging in other educational activities at their own desks during this time.

**Materials**

**Preschool Language Scales - 4th Edition (PLS4).** The PLS4 has been described in detail in Chapter 6 (please refer to p. 94). This standardized assessment measures language skills in children from birth to 6 years 11 months. However, this age norm is based on a typically developing sample. While the participants in this study ranged from 7-11 years, significant communication impairments are central to an Autism Diagnosis and hence the researchers considered this an appropriate language assessment to utilise.

**Non-arbitrary Same Different Relational Evaluation Protocol (NSD-REP) Procedure.** A description of this protocol was provided in the previous chapter (please refer to pp. 94-95).

In addition to these measures, a token economy was also used. This system was utilised given that it was a part of the pre-experimental assessment/training environment the participants were familiar with. Hence, removal of the token economy, especially with respect to a novel procedure may have resulted in difficulties with on-task behaviour for the children involved. For the PLS4
assessment, tokens (which were exchangeable for preferred tangible items/activities) were provided for appropriate sitting and listening behaviour. During the intervention phase, a single token was provided for each correct response on the NSD-REP. No tokens were provided for incorrect responding.

**Pre-assessment.** Before exposure to the core assessment measures, each participant was exposed to a brief test of their textual behaviour with respect to the textual stimuli “Yes”, “No”, “Same” and “Different”, each of which were employed as antecedents in the NSD-REP protocol. More specifically, the textual stimuli were presented individually a number of times in quasi-random order and participants were required to produce the correct vocal response for each. Participants were also required to produce correct textual behaviour with respect to these stimuli when they appeared onscreen at the start of computer-based assessment and training sessions.

**Assessment with PLS4.** Duration of testing ranged from 60 to 90 minutes; however, participants received five minute breaks every 15 minutes. Feedback for answering itself was not provided. However, sitting in one’s seat and listening to the researcher was reinforced with social and tangible reinforcement, namely, praise and tokens exchangeable for a preferred reinforcer, respectively. The PLS4 assessment was carried out as per the specifications of the manual. That is, the primary researcher presented questions from the manual orally and participants were expected to respond orally or indicate their answer using hand gestures (for example, pointing to pictures or objects). Questions were repeated verbatim if no response was given after 10 seconds. If no response was given after 20 seconds, an incorrect score was recorded. The assessment was terminated when participants scored incorrectly on seven consecutive questions, or when they had completed the assessment.
**NSD-REP Assessment.** After PLS4 testing sessions were completed, sessions of NSD-REP assessment started. Testing in this case lasted between six and seven minutes in total per person. This protocol assessed the evaluation of non-arbitrary same and different relations at three levels of complexity: (i) “Same” and “Different”; (ii) “Yes” and “No” and; (iii) “Yes” and “No” with “Not”. Participants were instructed that they were going to play a computer game in which they had to answer questions by pressing on one of two screen locations. If the participant showed signs of boredom or tiredness they were given a break and work was resumed after one minute. Participants were praised for sitting appropriately and listening to the question on each trial. If participants asked the researcher for an answer, they were told to try their best to answer unaided. In the case of all children, the NSD-REP assessment was completed in one session. The average time to complete the protocol was between five and ten minutes.

**NSD-REP Training Intervention.** Three participants (P1, P2 and P3) who passed Level 1 but failed Level 2 were chosen for a training intervention in a single-subject concurrent MBD across participants (Baer, Wolf, & Risley, 1968). The dependent variable or target behavior in this case was the percentage of correct Level 2 responding. Successful completion was attaining 90% correct on two consecutive sessions.

Sessions (involving either baseline or training intervention, depending on the phase in operation for a particular participant) were conducted once daily and consisted of 20 trials. The order in which participants received their sessions on a particular day was randomized across all three. Prior to a session, the instructor asked the participant if they would like to play the same and different game. If they agreed, they were brought to the table where they chose a reinforcing item or activity.
that they wanted access to on completion of that day’s session. They were then instructed to press the start button when ready to start.

During the baseline phase, participants were exposed to Level 2 of the NSD-REP with no feedback provided for correct or incorrect responses. After three baseline sessions, Participant one commenced the training intervention. Subsequent to a correct response, the computer presented the auditory stimulus “Correct, well done” accompanied by a visual display consisting of a grid of coloured stimuli including stars, small animals and shapes. After this, the screen cleared for 0.5 seconds before the start of the next trial. Correct responding was also reinforced with praise and tokens from the experimenter. Tokens for correct responding were stuck to a token board that displayed an image of the item or activity to which the participant wanted access. For incorrect responses, the computer presented the auditory stimulus “Wrong” with no visual accompaniment and the program automatically stopped to allow the researcher to provide feedback for incorrect responding. This was achieved by repeating the auditory (question) prompt and providing a further vocal and gestural prompt to highlight the correct answer.

During the training intervention, a stimulus set was used for two consecutive sessions before a novel stimulus set was introduced. After P1 had reached criterion (i.e., 90% on two consecutive sessions), the training intervention began for P2. After P2 had reached the same criterion, the training intervention began for P3. P1 received three NSD-REP training sessions. P2 received four NSD-REP training sessions. P3 received five NSD-REP training sessions.

**Re-Assessment on NSD-REP and PLS4.** Training took place over the course of several months starting in January and ending in early May. Two weeks
after training ended (i.e., approximately mid-May), all three participants involved in training were re-tested on both the NSD-REP and PLS4.

Results

Assessment

Demographic information and scores for the NSD-REP and PLS4 are provided in Table 7.1. NSD-REP scores were calculated as outlined in the Materials section in Chapter 6 (please see p. 92). Scores for the PLS4 (raw and age equivalent) were calculated as per the manual (Zimmermann et al., 2002).

Table 7.1.

**Demographic Information, Pre-Intervention NSD-REP and Pre-Intervention PLS4 Scores**

<table>
<thead>
<tr>
<th>Pt. No.</th>
<th>Age</th>
<th>Sex</th>
<th>NSD-REP L 1</th>
<th>NSD-REP L 2</th>
<th>NSD-REP L 3</th>
<th>NSD-REP T</th>
<th>PLS4 T</th>
<th>A E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7:0</td>
<td>M</td>
<td>10 (P)</td>
<td>0 (F)</td>
<td>0 (F)</td>
<td>10</td>
<td>120</td>
<td>5.02</td>
</tr>
<tr>
<td>2</td>
<td>7:3</td>
<td>M</td>
<td>10 (P)</td>
<td>5 (F)</td>
<td>0 (F)</td>
<td>15</td>
<td>111</td>
<td>4.04</td>
</tr>
<tr>
<td>3</td>
<td>9:4</td>
<td>M</td>
<td>10 (P)</td>
<td>1 (F)</td>
<td>4 (F)</td>
<td>15</td>
<td>120</td>
<td>5.02</td>
</tr>
<tr>
<td>4</td>
<td>11:0</td>
<td>M</td>
<td>1 (F)</td>
<td>1 (F)</td>
<td>0 (F)</td>
<td>2</td>
<td>93</td>
<td>3.03</td>
</tr>
<tr>
<td>5</td>
<td>8:4</td>
<td>M</td>
<td>2 (F)</td>
<td>1 (F)</td>
<td>0 (F)</td>
<td>3</td>
<td>84</td>
<td>2.10</td>
</tr>
<tr>
<td>6</td>
<td>9:1</td>
<td>M</td>
<td>2 (F)</td>
<td>0 (F)</td>
<td>0 (F)</td>
<td>2</td>
<td>79</td>
<td>2.07</td>
</tr>
<tr>
<td>7</td>
<td>9:3</td>
<td>M</td>
<td>0 (F)</td>
<td>1 (F)</td>
<td>0 (F)</td>
<td>1</td>
<td>63</td>
<td>1.01</td>
</tr>
<tr>
<td>8</td>
<td>7:5</td>
<td>M</td>
<td>2 (F)</td>
<td>2 (F)</td>
<td>0 (F)</td>
<td>1</td>
<td>75</td>
<td>2.05</td>
</tr>
<tr>
<td>9</td>
<td>10:6</td>
<td>F</td>
<td>0 (F)</td>
<td>3 (F)</td>
<td>2 (F)</td>
<td>5</td>
<td>72</td>
<td>2.04</td>
</tr>
</tbody>
</table>
A Spearman’s rank order correlation of $\rho = .82$ ($p = .0192$) was found between the NSD-REP and PLS4.

**Training Intervention**

Based on the results from their initial exposure to the NSD-REP protocol, three participants were selected for a training intervention in Level 2 responding using a concurrent MBD.
Study 6 Data

Figure 7.1 Study 6. Percentage of correct responses for Level 2 of the NSD-REP during phase I baseline (BL) and phase II intervention (INT) for each of the three participants involved. Intersection lines indicate progression to the next phase of the study. Numbers in brackets after participant numbers indicate number of that participant in the pre-assessment.

Figure 7.1 shows the outcome of the training intervention. The baseline responding for P1 remained relatively stable for three sessions (as may be seen there
was a slight upward trend after the first data point but none thereafter; nevertheless, this can be seen as a slight weakness in the context of an MB design). After intervention, scores for P1 improved rapidly and responding met criterion (90% on two consecutive sessions) within three sessions. Baseline responding for P2 and P3 remained stable during this time. After P1 met criterion, the intervention was introduced for P2. Responding for P2 met criterion within four sessions. Meantime, the baseline for P3 continued to be stable. After P2 had met criterion, the intervention was introduced for P3 and responding for this participant met criterion within five sessions. Table 7.2 shows ages and scores on the NSD-REP and PLS4 for P1-P3 both pre- and post-intervention.
Table 7.2.

Ages, scores on NSD-REP and score and age equivalence for PLS4 at pre- and post-intervention

<table>
<thead>
<tr>
<th>Pt. No.</th>
<th>Age (Y:M)</th>
<th>NSD-REP</th>
<th>PLS4</th>
<th>Age (Y:M)</th>
<th>NSD-REP</th>
<th>PLS4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>L1</td>
<td>L2</td>
<td>L3</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>1</td>
<td>7:0</td>
<td>10 (P)</td>
<td>0 (F)</td>
<td>0 (F)</td>
<td>10</td>
<td>120</td>
</tr>
<tr>
<td>2</td>
<td>7:3</td>
<td>10 (P)</td>
<td>5 (F)</td>
<td>0 (F)</td>
<td>15</td>
<td>111</td>
</tr>
<tr>
<td>3</td>
<td>9:4</td>
<td>10 (P)</td>
<td>1 (F)</td>
<td>4 (F)</td>
<td>11</td>
<td>120</td>
</tr>
</tbody>
</table>

Note: NSD-REP = Non-arbitrary Same and Different Relational Evaluation Procedure; L1 = Level 1; L2 = Level 2; L3 = Level 3; T = Total P = Pass; F = Fail; PLS4 = Pre-School Language Scale 4th Edition; T = Total Language Score; AE = Age Equivalent.
Discussion

Study 6 aimed to extend the work presented in Chapter 6, as well as REP work in the literature (e.g., Cassidy et al., 2011; Kilroe et al., 2014), by investigating the utility of the NSD-REP as an assessment and training tool in children with autism. To commence, nine children were assessed on the NSD-REP, whose core focus was contextual control over non-arbitrary same and difference relations, as well as on the PLS4, a standardized test of language. The results demonstrated a significant correlation between the two.

Subsequently, three children who passed NSD-REP Level 1 but failed Level 2 were trained on the latter using a concurrent MBD. All three participants met criterion within a number of sessions of training (i.e., three, four and five respectively for Participants 1, 2 and 3). This pattern of findings is similar to those found in Study 3, where typically developing children took between three and seven sessions to attain criterion. An additional link between these studies is that both cohorts of participants had similar age equivalent PLS4 scores at pre-test. For instance, mean AE scores for children in Study 3 were approximately four years and four months, comparable to scores in this study, which were four years and ten months. This provides further evidence that children with language scores of between four and five can respond successfully to training on Level 2 of the NSD-REP which involves relatively complex control over relational responding and arguably also, as discussed in an earlier chapter, represents relatively abstract relational responding. Moreover, the post-intervention results provide additional support for the importance of relational responding for human language and suggest the potential efficacy of the REP as a relational assessment/training format and of the NSD-REP as a form of REP for young children with ASD.
One aspect of the data not highlighted in the results section was the fact that improvements on the NSD-REP, based on training of Stage 2, were accompanied in the case of all three participants by improvements in PLS4 age-equivalent scores. Specifically, P1 showed an improvement of six months, P2 of eleven months and P3 of fifteen months in age-equivalent performance on this measure. It might be derived from this pattern that NSD-REP training may have been at least partially responsible for these improvements, however, the results need to be interpreted with caution for a number of reasons.

First, there was no control condition with which to compare this pattern. Second, it was based on the training of only three participants. Third, it is possible that the time elapsed between PLS4 test and re-test was insufficiently long and therefore some type of carryover effect may have at least contributed to the rise seen. Apart from this, there are theoretical reasons for scepticism that these improvements might have been largely a function of the NSD-REP training. These include the fact that the scope of training in terms of the relational skills actually trained (i.e., a combination of “Yes/No” and “Same/Different” contextually controlled non-arbitrary relations) and the quantity of training to which participants were exposed (i.e., each received relatively few actual training sessions) would not be expected to have a substantial effect on language skills, in the relatively broad sense, as measured by the PLS4. On the other hand, these results are at least intriguing and should prompt further better controlled research (e.g., using larger sample sizes, a control group and randomised assignment to condition) with similar types of population. Such studies might examine both the effects of relatively simple relational interventions (similar to that used in the current study) as well as of more
elaborate comprehensive relational interventions (e.g., including a range of both non-
arbitrary and/or arbitrary relations).

A number of related points can also be made in this respect. For instance, while this study did consider the effect of the training intervention on intellectual potential, it did not examine other possible forms of generalization such as to linguistic tasks involving same and different relations. This too might naturally come under the remit of a more substantial training and testing study. In addition, unlike Studies 2–4 presented in Chapter 6, the re-administration of baseline and follow-up data were not taken subsequent to training due to lack of access to participants as a result of school holidays (which was not an issue in the daycare facilities in which previous studies were conducted). Hence, it is not clear whether this responding would be maintained without feedback and or after a time period had elapsed. A further limitation of the current study was the absence of a standardised cognitive assessment (i.e., the Stanford Binet fifth edition or SB5), which had correlated with the NSD-REP in Study 1. Intellectual potential as well as language related tasks would have been preferable in the context of this population. However, the SB5 was unavailable at the time of testing and given the limited time frame for assessment and training, the PLS4 was utilised alone. Further, there was a limited amount of relational responding assessed and trained using this REP protocol, namely non-arbitrary same and different relations.

However, while these data might indeed prompt speculation as to the potential of REP-based assessment and training of relations themselves, Study 6 was not focused on the use of the REP to assess/train a particular pattern of relational responding in children with ASD. Rather, it sought to investigate the potential of the REP as a format for assessment/training in such a population. That is, non-arbitrary
same and difference relations themselves were not the target in this case, rather, the objective was appropriate REP type contextual control over these relations (for a more detailed explanation please refer to the General Discussion in Chapter 6). Nonetheless, although the results from Study 6 are tentative, they provide some preliminary evidence for utility of the NSD-REP as an educational instrument for children with ASD.

**Summary**

To conclude, the current study provides evidence for the important contribution of the REP in assessing and training relational responding in individuals with ASD. This study successfully demonstrated how relational responding can be established using the NSD-REP. Consistent with previous findings (including the studies presented in Chapter 6), the intervention successfully trained affirmation and disconfirmation as a generalized operant with this cohort of participants, and appeared to have a positive effect on language ability, though with previously provided caveats.

This research adds to the few examples of relational skills training with individuals with ASD in the literature (see McLay et al., 2013 for review). The results of the current study appear to be significant with regard to training of children with ASD, at least those with age equivalent scores on standardised language tests of four years of age or higher. Although it is impossible to generalize across children from any given population, the results of this intervention did generalize across three participants with somewhat similar language and intellectual profiles. Hence, this research arguably adds to our knowledge with respect to conducting future research teaching relational responding in children with ASD. Furthermore, the current findings further encourage research using the NSD-REP protocol, to
teach children with and without developmental difficulties, the skills required to respond successfully to the various levels of the protocol.
Chapter 8

Studies 7, 8 & 9

Assessing and Training Relational Evaluation in Mid-Childhood using the NSD-REP*

* The material in this chapter is in preparation for journal submission as: Hayes, J., & Stewart, I. (in preparation). Assessing and Training Complex (Negation Controlled) Responding in Mid-Childhood using the NSD-REP (Non-Arbitrary Same and Different Relational Evaluation Procedure.
In chapters 6 and 7, a series of experiential investigations indicated that: (i) performance on an REP-based protocol (the NSD-REP), specifically involving evaluation of non-arbitrary same/different relations, correlated with linguistic and intellectual potential more generally in young (2-4 year old) typically developing children; and (ii) appropriate REP performance could be effectively trained in several age groups of young typically developing children, as well as in children with ASD.

The success of these studies reaffirms the importance of generalized relational responding with respect to intellectual potential, including linguistic and cognitive skills. It also provides solid evidence that the NSD-REP as a methodology can be used to assess and train relational responding in young children both with and without developmental delay. In order to provide further verification of the NSD-REP as a potential assessment and training instrument, as well as to allow some measure of comparison across differing age groups, one potential domain for extension of this work is with slightly older children (i.e., aged 6-7 years). This chapter commences with a brief rationale for the current series of studies (7-9). Subsequently, the research methods are presented and the results and their implications are detailed.

Rationale for Studies 7-9

A critical feature of REP-type procedures is that they enable testing of relatively complex relational responding. For instance, the Implicit Relational Assessment Procedure (IRAP) typically requires that participants should respond flexibly to varying levels of negation/opposition contextual control across trials. For example, in Vahey, D. Barnes-Holmes, Y. Barnes-Holmes, and Stewart (2009), an IRAP protocol examining relational responding with regard to self-worth required
participants to relate either their own name (e.g., “Nigel”) or their name with “not” in front of it (e.g., “not Nigel”) as being either “Similar” or “Opposite” to either “Good” adjectives or “Bad” adjectives. Furthermore, on some blocks of trials they had to respond as if they were good and others were bad (e.g., by choosing “Nigel” similar to “Good” or by choosing “not Nigel” opposite to “Good”) while on other blocks of trials they had to respond as if they were bad and others were good (e.g., by choosing “Nigel” similar to “Bad” or by choosing “not Nigel” opposite to “Bad”). Hence, in this IRAP, participants had to respond flexibly along three different dimensions of contextual control i.e., presence versus absence of “Not”, similar versus opposite, and pro- or anti-self bias. The requirement to engage in this type of complex pattern of relational responding is not atypical of the IRAP where it has been shown to be useful, not simply for investigating participants’ idiosyncratic relational history but also for assessing intellectual potential (e.g., O’Toole & Barnes-Holmes, 2009).

As previously stated, a core aim of the current work is to examine the capacity of children to be assessed and trained using the REP. Responding in accordance with complex patterns of relational responding and in particular various combinations of negation relations is a common feature of prominent applications of the REP, such as the IRAP. Thus enabling testing of this kind is a particular advantage of the REP more generally. As such, it was deemed appropriate to test children’s capacity to show or to learn to engage in relatively more complex relational responding and in particular that involving multiple negation. Level 3 of the NSD-REP was designed to test relatively complex relational responding involving negation and a focus of the work presented in the current chapter is to
examine assessment and training within this level. Before going into the details of these studies, negation itself will be briefly discussed.

**Negation**

Within the mainstream psychological literature, negation is regarded as a critical component of human language and cognition. It is seen as central to our conceptualisations of the physical and social worlds, laying the necessary foundations for hypotheses and science by enabling us to posit what is not the case (Altman, 1967). Negative words and morphemes like “no,” “not,” and “-n’t” emerge early in children’s productive speech, and are described as enabling speakers to express multiple, related ideas (Nordmeyer & Frank, 2014). Despite its early emergence, it has been found that adults and children alike, experience difficulty with the nuances of negation (e.g., Clark & Chase, 1972; Kaup, Ludtke, & Zwaan, 2006; Nordmeyer & Frank, 2014).

In the language of RFT, negation can be conceptualised as a form of contextual control which allows a very important level of influence in relation to the environment. From this point of view, its emergence is seen as dependent on multiple exemplar training (MET) in which stimuli such as “no,” “not,” and “-n’t” come to be established as cues for this particular form of framing, which is similar in important respects to “Difference” framing. While there has been some RFT work conducted on the latter, this work has been mainly conducted with adults (e.g., O’Hora, Pelaez, & Barnes-Holmes, 2005). In relation to negation cues specifically, there is, at the current time, relatively little RFT research work with either adults or children (though see O’Connor, Rafferty, Barnes-Holmes, and Barnes-Holmes, 2009 as one possible exception). As such, we must rely on more mainstream work to provide clues as to its emergence and characteristic patterns.
Within the psycholinguistic and psychological literature, negation effects with adults have most frequently been explored in verification paradigms using true and false un-negated (affirmative) and true and false negated expressions. The following are examples of the four statements: (i) True un-negated (affirmative): “Four is an even number”; (ii) False un-negated (affirmative): “Six is an odd number”; (iii) True negated: “Four is not an odd number”; and (iv) False negated: “Six is not an even number” (Evans, 1982).

Employing true and false un-negated and negated expressions, the results of several investigations have illustrated delayed comprehension of negative sentences compared to positive (or un-negated) sentences (Carpenter & Just, 1975; Clark & Chase, 1972; Just & Carpenter, 1971; Kaup et al., 2006). As illustration, Kaup and colleagues (2006) asked participants to read sentences such as “The door was open” or “The door was not closed,” and were asked to name a picture that either matched (an open door) or did not match (a closed door) the sentence after a variable delay. When the picture probe followed an un-negated or affirmative sentence, participants were faster to label the matching probe than the mismatching probe after a 750ms delay; however, this effect did not emerge with negative sentences until a 1000ms delay. From a behaviour analytic perspective, these results suggest perhaps that the aforementioned participants had received insufficient MET opportunities to establish “no,” “not,” and “-n’t” as cues for this form of relational framing. Hence, they demonstrated an inability to respond to negation based sentences as rapidly as un-negated sentences, in which they would have received adequate MET through natural environmental contingencies. It is also possible that the delay may have been at least partially a function of the additional complexity of the contextual control
involved; younger children may be less likely to come under appropriate control due
either to less experience of language in general or of complexity.

With reference to children, mainstream research on negation is predominately
concerned with acquisition, specifically on the production (e.g., Bloom, 1970, 1993;
Pea, 1980; Cameron-Faulkner, Lieven, & Theakston, 2007) or interpretation of
negated sentences (de Villiers & Tager Flusberg, 1975; Nordmeyer & Frank, 2014;
Wojtecka, Koch, Grimm, & Schulz, 2011). For instance, assessing comprehension
of negation, Nordmeyer and Frank (2014) collected eye-tracking data while children
(age 2-5 years) watched a video in which they heard positive and negative sentences.
Negative sentences, such as “look at the boy with no apples,” referred to a boy with
nothing or a boy with an alternative object (e.g., a present). All children showed
greater difficulty in resolving the referent when negative sentences referred to the
boy with nothing. In addition, three and four year-old children showed an initial
tendency to look away from the target and towards the named noun when the
referent of the negative utterance was an alternative object. The results suggest that
even though children begin to come under influence of negation cues by
approximately 12 months (Bloom, 1970; Pea, 1980), appropriate levels of contextual
control are not as quickly or readily established as might be expected or desired.

At the same time, there appears to be little or no research on how best to
establish such appropriate control. Because negation is a fundamental element of
language and cognition, it seems important to begin to explore this question. While
the current work is not focused specifically on the emergence or training of negation
in children per se, it can perhaps contribute to increased information concerning the
utility of the REP as a method for doing so.
Age Groups Employed

The assessment and training that constituted the work reported in Chapter 6 focused on relatively young children between the ages of 2-4 years. For the purposes of the studies reported in the current chapter, it was decided instead, at least in the initial stages, to focus on an older group, more specifically, children aged 6-7 years old (i.e., middle childhood). For one thing, this would enable comparison between performances on the NSD-REP by more than one age group. Given the success of the studies reported in the previous chapters, which involved relatively young children, one question that might be asked is whether the NSD-REP protocol might equally successfully be deployed with a slightly older group. For example, might this protocol correlate with intellectual performance in 6-7 year olds and might it be used to train relational responding in this group if they showed deficits?

In addition, the older group might more readily allow investigation of Level 3 of the protocol. The performances seen in the previous study, involving the younger children, wherein none of the participants could pass Level 2 until they were trained, and even those who were trained on the latter showed difficulty on Level 3, suggested that children in this age range might not be the best population with respect to initial investigation of ‘not’ control (negation) on Level 3. In contrast, it seemed possible that the 6-7 year olds would reach Level 3 without training. Then perhaps at least some of them would fail at this point and thus training might be employed at this level. One question concerned how well this 6-7 year old group might do in this regard. Might training be successfully applied and might we be able to demonstrate this in a controlled, empirical fashion? Also, assuming that older children could be trained to pass Level 3, a further question concerned whether, younger children might be trained to pass it also.
Current Studies

The studies reported here focused on: (i) assessment / screening (Study 7); and (ii) training (Studies 8 and 9) of NSD-REP performance, with a focus on Level 3 in particular, in children ranging from 4-7 years of age. In Study 7, the NSD-REP protocol was correlated with a variety of assessments of intellectual ability in children aged 6-7 years. The correlated assessments included the Relational Abilities Test (RAI; Cassidy, Roche, & Hayes, 2011) which was used as a measure of relational framing ability, as well as standardised assessments of intellectual ability (i.e., the Stanford-Binet, 5th Edition; SB5) and of educational attainment (i.e., the DPRT-R [Drumcondra Primary Reading Test-Revised] and DPMT-R [Drumcondra Primary Math Test-Revised] tests). A standardised language assessment was not included in the current study as the Drumcondra Reading Test (see description below) was deemed a more suitable test of language in the context of this particular age group.

The rationale for Study 7 included: (i) to investigate if a relatively simple NSD-REP protocol targeted at relational ability might correlate with more complex / advanced relational and intellectual measures in 6-7 year olds; (ii) to compare performance on the NSD-REP protocol of slightly older children with that seen in the younger groups in the previous studies; and (iii) to identify suitable participants for subsequent training studies. In Study 8, MET was used to teach children aged seven years, criterion level responding on Level 3 of the NSD-REP protocol. Similarly, Study 9 utilised a refined and adapted NSD-REP protocol to teach Level 3 responding to a younger cohort of children (i.e., about four years old). These latter studies involved training appropriate responding in the context of relatively complex relational responding involving multiple negation cues.
Study 7: Method

Participants

Twenty three children \((M = 7.05, SD = .29; 12\) female), aged between six and seven years old, participated. They were recruited from two primary schools in the suburbs of Limerick city, Ireland. The first school is state-funded and has a student population of almost eight hundred (87% Irish; 90% Catholic) and 40 full time teaching staff. The second school is also state-funded and is one of 44 multi-denominational schools in Ireland, with 124 students (78% Irish) and nine permanent teaching staff. All participants were in first class. Recruitment materials were given to the caregivers of children by the respective class teachers. Children were included based both on written parental consent (i.e., as signified by a signed form returned to the class teacher) and their own verbal assent. Participants were informed that they could withdraw from the study at any time. All of the children enrolled in the study were Caucasian and none were receiving special education services in reading or mathematics.

Materials

**Stanford Binet Intelligence Scales 5th Edition (SB5) and the NSD-REP**

**Computer Protocol.** The SB5 and the NSD-REP have been described elsewhere in this thesis (please see pp. 93 and 94-95, respectively).

**Irish standardised measure of reading ability.** The DPRT-R (Drumcondra Reading) is a group-administered, standardised test of achievement in reading, designed for pupils in Irish primary schools. There are ‘reading vocabulary’ and ‘reading comprehension’ sub-tests at all six levels. At each level, it is possible to generate a total reading score, based on a pupil’s combined scores on the reading vocabulary and reading comprehension sub-tests, as well as subtest scores. All
questions involve a multiple-choice format. Performance is reported with reference to raw scores (the number of correct answers marked), standard scores, STEN (Standard Ten) scores, and percentile ranks. In addition to norms for each level/form of the test, there is a test-wide scale, which can be used to monitor the development of an individual student over several years.

An estimate of the internal consistency of both the DPRT-R and DPMT-R (see next sub-section) has been obtained based on the Kuder-Richardson Formula (KR20). KR20 coefficients give a measure of the extent to which the questions in each test constitute a whole, reflecting the degree to which different parts of the test or different questions are measuring the same attribute. A KR20 coefficient of 0.93 has been obtained for the DPRT-R (Educational Research Centre, 2007).

**Irish standardised measure of mathematical ability.** The DPMT-R (Drumcondra Maths) is a group-administered, standardised test of achievement in mathematics, designed for pupils in Irish primary schools. The content is based on the 1999 Primary School Mathematics Curriculum. About 50% of the test items follow a multiple-choice format. The remainder are short-answer items, in which the pupil may be asked to write an answer to a question, or draw a diagram. Raw scores can be transformed to standard scores, STEN scores and percentile ranks. A KR20 coefficient of 0.93 has been obtained for the DPMT-R (Educational Research Centre, 2006).

The rationale for including the DPRT-R and the DPMT-R in this study was to assess whether the NSD-REP, which had correlated with cognitive and language measures in previous studies would correlate with Irish standardised tests of educational attainment. This would further support its utility as a potential educational tool.
**Relational Abilities Index (RAI).** The Relational Abilities Index (RAI) is an online assessment available at the website [www.raiseyouriq.com](http://www.raiseyouriq.com). For the purposes of this assessment, each participant was provided with a temporary account on the website and the test was taken on Apple iPad tablets to which the children were given access. The RAI assessment consists of onscreen statements and questions to assess the fluency of “Same/Opposite” and “More Than/Less Than” relational framing. There are 55 test trials during each of which the respondent is required to answer a relational task within a limited time (see Procedure for further details). The “Same/Opposite” relation is assessed on the first 28 trials while the “More Than/Less Than” relation is assessed on the remaining 27 trials. The RAI software records responses and after all test trials are completed, the total number of correct responses appears onscreen as immediate feedback and remains available for a prescribed period in a database for access by the researcher. The purpose of including this measure was to ascertain whether the NSD-REP would correlate with a more established REP protocol that assesses both non-arbitrary and arbitrary relational responding.

**Procedure**

All assessments given to each participant except for the Drumcondra tests were conducted in a single session in a quiet, empty classroom, adjacent to, or in close proximity to the general education classroom in the participant’s respective school. All 23 participants were assessed on an individual basis in this room on several measures described in the previous section including: (i) the SB5; (ii) the NSD-REP; and (iii) the RAI. These tests were administered in the same order for each child and testing sessions took approximately 55 minutes in total. No feedback was provided at any stage during these sessions.
SB5. This test was administered by the experimenter in one sitting according to standardised test instructions and took approximately 20 minutes on average to complete.

NSD-REP. After the SB5, the NSD-REP was presented. As described previously, this assessed the evaluation of non-arbitrary same and different relations at three levels of complexity including: (i) same and different; (ii) yes and no; and (iii) yes and no with not, and progression from one level to the next depended on meeting criterion for the preceding level. In the case of each level, first instructions were provided by the experimenter and then trials began. While trials were being presented, the researcher sat quietly beside the child and did not engage with him or her until the level had ended. If after exposure to a level a participant had succeeded in demonstrating a pre-determined percentage of correct responses (i.e., 80%) for that level, then they were deemed as having passed and were exposed to the next level. If he or she did not attain this target they were deemed to have failed that level and the assessment ended. In the case of all children, this assessment was completed in one session. The average time to complete the protocol was six minutes.

RAI. This was the final researcher-delivered assessment. At the start, the researcher carefully explained to the child what was required. Participants were told that they would be required to answer a series of questions; that for each question, statements containing nonsense words (e.g., VAC) would appear on the screen first, followed by the question; that in order to answer the question, they would have to read the statements carefully, and then click either “Yes” or “No” in response; that they only had 30 seconds to answer each question; and that they had to work on their own and answer each question in their head (i.e., without writing anything down). These basic instructions (see Appendix B) were also presented onscreen.
The assessment involved 55 trials. “Same/Opposite” relations were assessed in the first 28 trials while “More Than/Less Than” relations were assessed in the remaining 27. On each trial, participants would see one to three statements presented onscreen in addition to a question which was presented under the statements (see Figure 8.1). The response options “Yes” and “No” also appeared in the bottom right and left of the computer screen (with their positions counterbalanced across trials). For example, one possible task might have presented the statement, “FOC is opposite to WIS” on the top line, the statement “WIS is opposite to NIL” on line 2, the statement “NIL is the same as VAQ” on line 3 and the question “Is FOC the same as VAQ?” on the bottom. In this case, as FOC is opposite to WIS and WIS is opposite to NIL then FOC and NIL are the same; as NIL is the same as VAQ then FOC and VAQ are also the same; thus the correct response would be “Yes”. No feedback was provided on any trial. Instead, when all 55 trials had been presented, the participant was given an onscreen message telling them that the test was over. This assessment took approximately 28 minutes to complete.

Figure 8.1. Example of a Relational Abilities Index (RAI) trial.
DPRT-R (Drumcondra Reading) and DPMT-R (Math) Assessments.
The DPRT-R and DPMT-R assessments were administered to both groups by the
class teacher in the beginning of June in the year of the current study. The
assessments were given in the general education classroom in two separate sessions,
three days apart. Participants first completed the DPRT-R, which lasted
approximately 120 minutes. They subsequently completed the DPMT-R which
lasted 90 minutes. Participants completed Level 1 in both tests. This required them
to listen carefully to their teacher as she read out the test items, after which they were
required to record their answers in the test booklets. No feedback was provided
during testing. Both standardised measures were scored manually by teachers using
the answer keys in the test manuals. The results were reported with reference to raw
scores (the number of correct answers marked), standard scores, STEN scores, and
percentile ranks.

Results and Discussion
Scoring Protocols
Demographic information and scores for the SB5, NSD-REP, RAI and
Drumcondra assessments are provided in Table 8.1. Scores for the NSD-REP were
calculated as outlined in the NSD-REP assessment section. Scores for the SB5
(verbal and non-verbal), were calculated as per the SB5 manual guidelines (Roid,
2003). Scores for the RAI were calculated as the total number of trials correct out of
a maximum score of 55. Scores for DPRT-R (Drumcondra Reading) and the
DPMT-R (Drumcondra Math) were calculated as per the scoring key or stencil in
the test manuals.
Table 8.1

Demographic Information, Scores for NSD-REP, RAI, SB5 & Drumcondra Assessments

<table>
<thead>
<tr>
<th>Pt. No.</th>
<th>Age (Y:M)</th>
<th>Sex</th>
<th>NSD-REP</th>
<th>RAI</th>
<th>SB5</th>
<th>DPMT-R</th>
<th>DPRT-R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>L1</td>
<td>L2</td>
<td>L3</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>1</td>
<td>7:04</td>
<td>M</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>7:07</td>
<td>M</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>7:03</td>
<td>F</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>7:09</td>
<td>M</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>27</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>7:04</td>
<td>M</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>7:11</td>
<td>F</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>7:04</td>
<td>M</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>29</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>7:09</td>
<td>F</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>27</td>
<td>31</td>
</tr>
<tr>
<td>9</td>
<td>7:06</td>
<td>F</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>7:02</td>
<td>M</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>11</td>
<td>7:00</td>
<td>M</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td>19</td>
<td>31</td>
</tr>
<tr>
<td>12</td>
<td>7:01</td>
<td>M</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>7:11</td>
<td>F</td>
<td>9</td>
<td>8</td>
<td>3</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>14</td>
<td>7:00</td>
<td>F</td>
<td>9</td>
<td>10</td>
<td>6</td>
<td>25</td>
<td>29</td>
</tr>
<tr>
<td>15</td>
<td>7:02</td>
<td>F</td>
<td>10</td>
<td>9</td>
<td>2</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>16</td>
<td>6:10</td>
<td>M</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>17</td>
<td>7:10</td>
<td>F</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td>19</td>
<td>33</td>
</tr>
<tr>
<td>18</td>
<td>8:00</td>
<td>M</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>19</td>
<td>7:11</td>
<td>M</td>
<td>9</td>
<td>10</td>
<td>6</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>20</td>
<td>7:06</td>
<td>F</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>21</td>
<td>7:00</td>
<td>F</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>22</td>
<td>7:10</td>
<td>F</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>26</td>
<td>34</td>
</tr>
<tr>
<td>23</td>
<td>7:11</td>
<td>F</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td>19</td>
<td>27</td>
</tr>
</tbody>
</table>
Note: NSD-REP=Non-Arbitrary Same and Different Relational Evaluation Procedure; L 1 = Level 1; L 2 = Level 2; L 3 = Level 3; RAI= Relational Abilities Index; SB5 = Stanford-Binet 5th Edition; T = Raw score for total (composite) scale; V = Verbal; NV = Nonverbal; AE = Age Equivalent; IQ = Intelligence Quotient; DPMT-R = Drumcondra Primary Maths Test- Revised; DRPT-R = Drumcondra Primary Reading Test- Revised.

Correlations Between the NSD-REP and Other Measures

Table 8.2 displays a matrix of Spearman rank order correlations among the experimental measures and, in the case of the SB5, its key sub-scales. This shows a correlation between the NSD-REP and SB5 Total (SB5 T) scores \( (r = 0.67; p < .0001) \) as well as between the NSD-REP and the two SB5 sub-scales. As regards the latter, the correlation with the non-verbal SB5 score was 0.54 \( (p < .05) \) while that with the verbal SB5 score was 0.49 \( (p < .05) \). The NSD-REP was also found to be correlated with the DPMT-R (Drumcondra Math) score \( (p = 0.54; p < .05) \), though not with the DPRT-R (Drumcondra Reading) score \( (p = 0.4; p > .05) \). Thus, briefly, the overall pattern of results shows that the NSD-REP correlated with both a measure of cognitive ability (i.e., the SB5) and it’s sub-scales as well as with a standardised school Math Assessment.

Correlations Amongst Alternative Measures

Table 8.2 also reveals correlations among the alternative measures themselves. The SB5 Total scale was correlated with both the Drumcondra Math \( (r = 0.51, p < .05) \) and Reading \( (r = 0.51, p < .05) \) assessments, while the SB5 verbal sub-scale was correlated with the Math score alone \( (r = 0.46, p < .05) \). Meantime, the RAI was correlated with none of the other measures.
Table 8.2

*Spearman’s rank order correlations (n = 23) between key assessment measures*

<table>
<thead>
<tr>
<th></th>
<th>NSD-REP</th>
<th>RAI</th>
<th>SB5-T</th>
<th>SB5-V</th>
<th>SB5-NV</th>
<th>DPMT-R</th>
<th>DPRT-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSD-REP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAI</td>
<td>-0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB5-T</td>
<td>0.67***</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB5-V</td>
<td>0.49*</td>
<td>0.07</td>
<td>0.73***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB5-NV</td>
<td>0.55**</td>
<td>0.05</td>
<td>0.67***</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPMT-R</td>
<td>0.54**</td>
<td>0.05</td>
<td>0.51**</td>
<td>0.46*</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPRT-R</td>
<td>0.4</td>
<td>0.1</td>
<td>0.51**</td>
<td>0.37</td>
<td>0.34</td>
<td>0.76***</td>
<td></td>
</tr>
</tbody>
</table>

Note: Asterisks indicate a statistically significant correlation (2-tailed); * p < .05.  ** p < .01.  *** p < .001.  NSD-REP = Non Arbitrary Same and Different Relational Evaluation Procedure; RAI = Relational Abilities Index; SB5 = Stanford-Binet 5th Edition; T = Raw score for total (composite) scale; V = Verbal; NV = Nonverbal; DPMT-R = Drumcondra Primary Maths Test-Revised; DRPT-R = Drumcondra Primary Reading Test-Revised.

The fact that the SB5 total score was correlated with both the DPMT-R (Drumcondra Maths) and DPRT-R (Drumcondra Reading) assessments might have been predicted given that all three are standardised measures, the first of cognitive ability and the latter of attainment in domains in which cognitive ability is important. Meantime, the fact that the RAI did not correlate with any of these measures is relatively unexpected, especially as relational ability as measured by the RAI is seen as critical to intellectual performance as conceptualised by RFT. How might this be explained? Anecdotal evidence in the form of verbal reports from many of the participants suggested that, despite careful explanation of the core requirements of the RAI relational task, both onscreen and by the experimenter, and apparent understanding on the part of the participants, the latter found it very difficult. This
may have been a factor of age. Previous work using the RAI has focused on children in the 8.5-12 age range bracket (i.e., Cassidy et al., 2011), while the current participants, at around 6-7 years of age, were younger than this. While, as suggested, RFT would suggest that derived relational responding provides the basis for intellectual performance and develops from an early age across the basic frames assessed in the RAI, there are other factors that may impinge upon performance of a particular population on a particular relational task. It may be that the range of derived relational ability assessed by the RAI is too high for children of this age. Alternatively, it may also be that completion of this test requires a degree of attention/focus that only develops after years of schooling. In any event, RAI scores seen in this study were relatively lower and showed much less variability in comparison to those seen in Cassidy et al. (2011), where the RAI was first used. The mean, standard deviation and range in the latter case were 46.75, 9.9 and 29-60.7 while the comparable figures in the current study were 27.43, 3.5 and 20-34. This pattern is likely a function of the fact that children in this age group have less experience of relations than children in older age groups such as those appearing in Cassidy et al. (2011). The resulting pattern and especially the lack of variability of the data may have been responsible for the lack of correlation with other measures.

In contrast with the RAI, the NSD-REP protocol that has been employed in this study correlated relatively well with other instruments including the SB5 and the DPMT-R (Drumcondra Math) assessment. Yet, the NSD-REP did not correlate with the RAI. Both the latter protocols are types of REP format. However, while the RAI tests for derived arbitrarily applicable relational responding ability from relatively simple to relatively complex, the NSD-REP differs. It is based ultimately only on contextually controlled non-arbitrary same/different relational responding, though
with varying degrees of contextual control over demonstration of that repertoire in turn required to perform well at different levels. Level 1 requires relatively straightforward responding in accordance with contextually controlled non-arbitrary same/different itself, Level 2 requires “Yes/No” contextual control and Level 3 requires both “Not” and “Yes/No” contextual control. The requirement for contextual control over relational responding may tap into a fundamental repertoire that is key to the learning of various patterns of derived relations over time. However, in contrast with the RAI, the NSD-REP utilises this repertoire without requiring familiarity with a variety of abstract forms of relational responding at various levels of complexity, as is the case with the RAI. As such, this might suggest that the NSD-REP would be a more reliable indicator / predictor of intellectual potential for a younger age group than the RAI.

One phenomenon which Level 3 of the NSD-REP taps into is of course negation. In the introduction, it was noted that in several studies there appears to be delayed comprehension of negative sentences compared to positive (or un-negated) sentences (e.g., Clark & Chase, 1972; Kaup et al., 2006). The results of the present program of research might be seen as showing similar results in a broad sense as many of the relatively older children involved in the current study performed well on Level 3 without any training. From Table 8.1, we can see that six children obtained the maximum score on Level 3 while a further eight scored at least six out of 10. In comparison, younger children in some of the previous studies performed extremely poorly on Level 3 (typically attaining zero) despite doing reasonably well on lower levels. As suggested in the introduction, perhaps younger children have received insufficient MET opportunities to come under appropriate negation-type control. It is also possible that their performance is a function of the additional complexity of
the contextual control involved in tests such as Level 3 of the NSD-REP; they may be less likely to come under appropriate control due either to less experience of language/relational responding in general or of complexity per se. These are issues which further exploration using the NSD-REP or similar methods with various age groups might investigate.

Apart from allowing comparison with alternative assessments and giving rise to theoretical issues with regard to development and training of particular repertoires, the outcome of the NSD-REP assessment in the current study also provided the basis for identifying individuals suitable for an experimental intervention. The first of two studies involving this work (Study 8) is reported next.

**Study 8**

Based on the results from the NSD-REP protocol used in Study 7, three participants who met criterion for Level 2 but not for Level 3 were randomly selected. The current study aimed to train these participants to meet criterion for Level 3 (“negation control over yes/no responding to non-arbitrary same/difference relations”) employing a concurrent multiple-baseline design (MBD) across participants.

**Method**

**Participants and Setting**

Three female participants, aged seven years old ($M = 7.06, SD = .06$), participated in this study. They were selected from amongst participants used in Study 7 who met criterion for Level 2 but not for Level 3 (i.e., participants 20, 23 and 21). All training sessions were conducted in a small empty classroom in close proximity to the participants’ general education classroom.
Materials

The NSD-REP protocol (Level 3) was presented on a touch screen Dell Latitude 2100 computer. Onscreen images (including, for example, animals, shapes, common objects and cartoon characters) were used as stimuli. These were organised into particular stimulus sets (named Stimulus Sets 1-5).

Target Behaviour and Experimental Design

The target behaviour for the current study was correct responding on Level 3 of the NSD-REP protocol. A single-subject non-concurrent multiple probe design across participants (Baer, Wolf, & Risley, 1968) was implemented. A multiple probe design is a variation of the multiple baseline design that features intermittent measures or probes during baseline. Experimental control is established when a targeted individual’s performance improves while the other participants behaviour remains stable. This design differed to previous studies in the thesis (i.e., Studies 2-4, 6) which utilised multiple baseline designs. The primary reasons for opting for this procedure were as follows: (i) based on the previous studies in this thesis a strong a priori assumption of stability could be made; and (ii) the researchers wanted to reduce baseline exposure to the intervention (i.e., Phase 1), given that participants in this study did not receive intervention until the previous participant met criterion (i.e., Phase 2) and maintained this responding in the absence of feedback one week later (i.e., Phase 3). Probes were taken every two to three sessions to ensure participants were not attaining criteria in the absence of training.

Progression of treatment from one phase to the next was based on certain pre-designated criteria being met (see the description of individual phases).
Procedure

On a typical testing day, one or more of the three participants was exposed to a session of testing or training, depending on the phase for each. A session was conducted in an empty classroom and consisted of 20 trials of testing or training. In cases in which more than one student was being exposed to a session, a random number generator was used to establish the sequence of participation.

Prior to the session for an individual participant, the experimenter first activated the NSD-REP protocol and ensured that the account for the participant about to partake in the ensuing session was open. Then, depending on what phase the participant was in, the experimenter turned the protocol feedback option on or off. Subsequently, the researcher went to the general education classroom and asked the participant in question if they would like to play the same and different game on the computer. If the participant agreed, they were brought to the classroom assigned for that session. The experimenter sat beside the participant throughout the session. In some phases, she would directly interact with her while in others there was no interaction (see below).

There were four phases in the experiment: (i) baseline; (ii) intervention; (iii) return to baseline; and (iv) follow-up. A random number generator was used to establish the order in which participants would receive the intervention.

Phase I-Baseline. During this phase, participants were exposed to sessions of unreinforced Level 3 trials. Each trial involved a new pair of stimuli quasi-randomly chosen from a designated stimulus set (i.e., Stimulus Set 1). On each trial the participant was required to respond to an auditory question concerning the sameness or difference of the non-arbitrarily related stimulus pair shown and containing the word “Not” (e.g., “Are these not the same?”) by touching either “Yes”
or “No” on the computer screen. No feedback was provided during baseline sessions. Progression from baseline to intervention happened for Participant 1 after three baseline probes. Progression for the other two participants happened a minimum of three baseline sessions after the last participant who had gone into intervention, assuming also that: there was a stable or downward trend in the data for the participant in baseline and; that the previous participant had achieved criterion for progression to the second baseline (no feedback) phase.

**Phase II-Intervention.** This phase differed from baseline in that: (i) novel stimulus sets (i.e., sets 3-5) were used; and (ii) auditory and visual feedback were provided for correct and incorrect responding. The automated feedback participants received was the auditory recording “Correct, well done” accompanied by visual imagery (stars and other icons) for correct responses and the auditory recording “Wrong” with no accompanying visual stimuli for incorrect responses. Subsequent to an incorrect response, the task on which the participant responded incorrectly was re-presented on the screen by the experimenter manually pressing a specific key. The experimenter then prompted the correct answer by vocally producing it (i.e., saying either “Yes” or “No”) and gesturing to the correct on-screen button. During the training phase, a new stimulus set (i.e., set 3) was introduced immediately on commencing and on every second session thereafter (drawing on sets 4-6 as needed). In addition, during this phase performance on stimulus set 1 was probed after every two to three training exposures as well as after criterion was reached. A participant progressed to the next (return to baseline) phase by showing 2 x 90% correct responding in two consecutive sessions. If, in contrast, there was a decreasing trend of three data points or if data remained stationary across three consecutive sessions,
it was intended that an alternative (remedial) phase would be introduced though this was not necessary.

**Phase III- Return to Baseline.** When a participant reached criterion in the intervention phase they were re-probed with the Phase 1 stimulus set (i.e., Set 1) both immediately as well as one week post-intervention. As in the case of the initial baseline, no feedback was provided during these probes.

**Phase IV- Follow up testing.** For this phase, which was carried out three weeks subsequent to the end of Phase III, testing of the target was conducted using a previously unseen stimulus set (Set 2).

**Results and Discussion**

Figure 8.2 displays the effects of multiple exemplar training (MET) of Level 3 responding (i.e., negation controlled affirmation/disconfirmation of non-arbitrary same/different relations) on percentage of correct responding as measured for three participants using a multiple probe design. MET produced an increase in the percentage of correct responses sufficient to allow all participants to meet criterion. Participant 1 demonstrated an immediate increase in percentage of correct responses on the introduction of the intervention in Ph II and required only three training sessions to meet criterion. These high scores were maintained when the baseline stimulus set was re-administered one week later (Ph III) and when a novel stimulus set was tested four weeks after training had finished (Ph IV). Participant 2 met the training criterion after six sessions. However, while they subsequently passed the immediate probe for Stimulus Set 1 responding (i.e., attained at least 90% correct), this participant failed to make criterion when the probe was re-administered one week later. Hence, the training phase was re-introduced and subsequent to two further sessions, criterion responding was shown both on training and on the probe
set and then both on the 1 week follow up of the probe (Ph III) as well as on the follow up session (Set 2, Ph IV). Participant 3 required five training sessions prior to demonstrating criterion performance. Criterion performance was then shown in the probe and maintained when the baseline stimulus set was readministered one week later (Ph III) and when a novel stimulus set was given three weeks after that (Ph IV).

These findings provide a controlled demonstration of MET as a means of training Level 3 NSD-REP performance. In doing so, they extend the findings from Studies 2, 3 and 6 in Chapters 6 and 7 respectively. They do so by providing further evidence of the utility of MET of relational responding. They also indicate that the effects of training are maintained several weeks later with new stimulus sets. Given the correlation with IQ and mathematical skill shown in Study 7, improving this repertoire is potentially important. Thus, results such as these, which show that it can be successfully trained and maintained with totally novel stimulus sets, are encouraging.
Study 8 Data

Figure 8.2. Study 8. Percentage of correct responses for Level 3 of the NSD-REP during phase I baseline (BL), phase II intervention (INT), phase III baseline (BL) and phase IV follow-up (FU), for each of the three participants involved.
All three participants in this study seemed to respond quite quickly to the intervention and thus it seemed particularly effective. This might also have been related to age of course as all three were seven year olds. Accordingly, a question arises as to the potential efficacy of providing training of this kind of repertoire to a slightly younger group. This is what the subsequent study attempted to examine.

**Study 9**

The aim of Study 9 was to extend the results of Study 8 by investigating whether a younger group of participants than those involved in the latter could be taught the same relatively complex pattern of relational responding, namely NSD-REP Level 3 responding (i.e., negation control over yes/no responding to non-arbitrary same / difference relations) in the context of a similar design, that is, a concurrent multiple probe design (MBD) across participants.

**Method**

**Participants and Setting**

Three children (2 female, 1 male), aged four years old ($M = 4.09, SD = .02$), participated in this study. They attended a private Montesorri school in Limerick, Ireland. Montessori schooling caters for children aged three to five years and utilises various behavioural techniques including antecedent (e.g., daily visual schedules) and consequence based strategies (e.g., group contingencies). These children were selected based on pre-screening which identified that they could successfully complete Level 2 but not Level 3 on the NSD-REP.

All three participants were involved in previous studies, namely Studies 1, 3 and 5, in which they were coded as participants 14, 12 and 21 respectively. Study 1, a correlational based study indicated that the children had age equivalent scores of 4.03 years or above in both the language (i.e., PLS4) and cognitive (i.e., SB5)
assessments. The research presented in the thesis thus far (Studies 1 and 6) has suggested that language age equivalence scores of 4 years predicts successful completion of Level 2 training and indeed this was shown in the case of each of the participants in the current study (Studies 3 and 6). The question under examination in Study 9 was whether children at this level might also be trained in the more complex (i.e., negation controlled) repertoire required for Level 3. An examination of pre to post-test differences was examined in Study 5, however, the post-assessment was not completed until the participants met criterion with Level 3 responding.

All testing and training sessions were conducted in a small room in close proximity to the participants’ general education classroom.

Materials

As for Study 8, the NSD-REP protocol was administered using a touch screen Dell Latitude 2100 computer. In addition, tangible items (e.g., stickers, erasers, sweets and colouring pages) were used as reinforcing consequences and an individualised score chart and pencil were used for each participant.

Target Behaviour and Experimental Design

The target behaviour (i.e., criterion level correct responding on Level 3 of the NSD-REP protocol) and the experimental design (i.e., non-concurrent multiple probe design) were the same as in Study 8.

Procedure

On a typical testing day, each of the three participants was run through one individual session. A session was conducted in an empty classroom and consisted of 20 trials of either testing or training, depending on the phase. A random number
generator was used to establish the daily sequence of participation for the three students.

Prior to the session for an individual participant, the experimenter first activated the NSD-REP protocol and ensured that the account for the participant about to partake in the ensuing session was open. Then, depending on what phase the participant was in, the experimenter turned the protocol feedback option on or off. Subsequently, the researcher went to the general education classroom and asked the participant in question if they would like to play the same and different game on the computer. If the participant agreed, they were brought to the classroom assigned for that session. The experimenter sat beside the participant throughout the session. In some phases, she would directly interact with her while in others there was no interaction (see below).

As indicated in the Participants and Setting section, the participants were selected based on pre-screening in which they passed Level 2 but failed Level 3 of the NSD-REP protocol. In addition, in initial pilot testing the participants were also exposed to multiple exemplar training of the target behaviour itself. As explained in the procedure section of previous studies there was a pre-established criterion for all training whereby if levels of correct responding showed a decreasing trend of three data points or if data remained stationary across three consecutive sessions, then the training would be considered unsuccessful and an alternative (remedial) phase would be introduced instead. For all three participants, direct multiple exemplar training of the target behaviour itself failed to establish the desired repertoire based on this criterion and thus the training was brought to an end and a remedial type of multi-phase training was instituted as the intervention instead. Please see Table 8.3 below.
Table 8.3

Baseline and Intervention Scores for Participants 1-3 using unadjusted NSD-REP Level 3

<table>
<thead>
<tr>
<th>Participant</th>
<th>Baseline</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (14)</td>
<td>50 50 50</td>
<td>50 50 50</td>
</tr>
<tr>
<td>2 (12)</td>
<td>50 50 55 55</td>
<td>50 55 45 45</td>
</tr>
<tr>
<td>3 (21)</td>
<td>55 50 50 55 55</td>
<td>50 55 50 50</td>
</tr>
</tbody>
</table>


Given that the participants were unable to respond to the unadjusted Level 3 of the NSD-REP, the researchers divided this level into three component elements in an effort to ascertain whether the additional phases would aid with the attainment of this skill. These adjusted phases included targeting “Yes” and “No” with same (a skill participants had acquired during the training of Level 2) and not same (a novel skill) relations initially. If participants demonstrated success with this phase of training, the next phase, which targeted “Yes” and “No” with different (established skill) and not different (a novel skill) relations. Again, if participants illustrated
success with this phase, “Yes” and “No” with not same and not different relations were targeted (which formed the unaltered Level 3). Hence, there were six phases in the current study altogether: (i) baseline; (ii) same/not same; (iii) different/not different; (iv) not same/not different; (v) return to baseline; and (vi) follow-up. A random number generator was used to establish the order in which participants would receive the intervention.

**Phase I - Baseline.** During this phase, participants were exposed to sessions of unreinforced Level 3 trials. Each trial involved a new pair of stimuli quasi-randomly chosen from a designated stimulus set (i.e., Stimulus Set 1). On each trial the participant was required to respond to an auditory question concerning the sameness or difference of the non-arbitrarily related stimulus pair shown and containing the word “Not” (e.g., ‘Are these not the same?’) by touching either “Yes” or “No” on the computer screen. No feedback was provided during baseline sessions. Progression from baseline to intervention happened for Participant 1 after three baseline probes. Progression for the other two participants happened a minimum of three baseline sessions after the last participant who had gone into intervention, assuming also that: (i) there was a stable or downward trend in the data for the participant in baseline; and (ii) that the previous participant had achieved criterion for progression to the second baseline (no feedback) phase (i.e., Phase V).

**Phase II - Training Same/Not Same.** Subsequent to three baseline sessions, Participant 1 commenced Phase II. Phase II, Phase III and Phase IV were all training phases. They trained same/not same relations, different/not different relations and not same/not different relations respectively. These phases all differed from baseline in that (i) novel stimulus sets (i.e., sets 3-5) were used and (ii) auditory and visual feedback were provided for correct and incorrect responding. The automated
feedback participants received was the auditory recording “Correct, well done”
accompanied by visual imagery (stars and other icons) for correct responses and the
auditory recording “Wrong” with no accompanying visual stimuli for incorrect
responses. Subsequent to an incorrect response, the task on which the participant
responded incorrectly was re-presented on the screen by the experimenter manually
pressing a specific key. The experimenter then prompted the correct answer by
vocally producing it (i.e., saying either “Yes” or “No”) and gesturing to the correct
on-screen button.

During the intervention phases (II, III, and IV), at the end of each session, the
experimenter displayed the participants score on their individualised score chart.
The rationale for this was twofold: (1) to maintain participants’ motivation levels
given the complexity of the skill being targeted; and (2) the children requested
feedback on their performance and self-monitoring behaviour has been linked to
improved work productivity and accuracy (e.g., Rock, 2005) as well as increased
academic engagement (e.g., Amato-Zech, Hoff, & Doepke, 2006) in young children.

Regardless of score, participants received social (i.e., praise) and tangible
(i.e., a sticker) reinforcement for performance. If, however, their scores had
increased since the previous training session, they also gained access to the lucky
bag. This reinforcement procedure differed slightly to those highlighted in Studies 2-
4, where the children were permitted to access the lucky bag at the end of every
training session, irrespective of performance.

Phase II trained same/not same responding. This meant that participants
were required to respond “Yes” or “No” to un-negated and negated “Same”
questions (i.e., “Are these the same/not the same?”) presented alongside two stimuli
(either same or different) on a trial. A stimulus set was used for two consecutive
sessions before a novel stimulus set was introduced. A new stimulus set (i.e., set 3) was introduced immediately on commencing Phase II and on every second session thereafter (drawing on sets 4-6 as needed). In addition, during this phase and the two subsequent training phases performance on stimulus set 1 was probed after every two to three training exposures as well as after a phase criterion was reached. A participant progressed to the subsequent phase by showing 2 x 90% correct responding in two consecutive sessions. If, in contrast, there was a decreasing trend of three data points or if data remained stationary across three consecutive sessions, it was intended that an alternative (remedial) phase would be introduced though this was not necessary.

**Phase III - Training Different/Not Different.** When participants reached criterion in the in Phase II (i.e., 90% across two consecutive sessions), they were re-probed with the Phase 1 stimulus set (i.e., Set 1) and progressed to Phase III. This was identical to Phase II with the exception that participants were required to respond to un-negated and negated non arbitrary “Different” relations (i.e., “Are these different/not different”) by choosing “Yes” or “No”.

**Phase IV - Training Not Same/Not Different.** When participants reached criterion in Phase III (i.e., 90% across two consecutive sessions), they were re-probed with the Phase 1 stimulus set (i.e., Set 1) and progressed to Phase IV. This phase was identical to Phase II and III with the exception that in this phase, participants were required to respond to negated “Same” and “Different” non-arbitrary relations (i.e., “Are these not same/not different”) by choosing “Yes” or “No”.

**Phase V - Return to Baseline.** When participants reached criterion in Phase IV (i.e., 90% across two consecutive sessions), they were re-probed with the Phase 1
stimulus set (i.e., Set 1), both immediately as well as one week post-intervention. As in the case of the initial baseline, no feedback was provided during these probes.

**Phase VI - Follow up testing.** For this phase, which was carried out three weeks subsequent to the end of Phase V, testing of the target was conducted using a previously unseen stimulus set (Set 2).

**Results and Discussion**

Figure 8.3 displays the effects of the multi-phase training of NSD-REP Level 3 (yes/no to “Not Same/Not Different”) responding on the percentage of correct responses in the context of a multiple probe design across participants. As revealed by the graph, the multi-phase training produced an increase in the percentage of correct responses sufficient to allow all participants meet criterion and to maintain accuracy in follow-up testing. All participants demonstrated an immediate increase in percentage of correct responses to un-negated and negated same (i.e., same/not the same) relations (Phase II). In Phase III, participants required between five and six training sessions to reach criterion in response to un-negated and negated different (i.e., different/not different) relations. Probes show that, despite reaching criterion on both Phases II and III, however, participants had still not improved on the target response. When negated same and different relations were combined in Phase IV thus training the target, the participants required between three and six training sessions to meet criterion. Subsequent to receiving this training of the combination / target pattern, participants also reached criterion on the probes for baseline (Stimulus Set 1) responding. After attaining criterion in Phase IV, baseline and follow-up/generalisation probes were administered one week and four weeks after Phase IV respectively. All three participants showed criterion levels of correct responding at both points.
Subsequent to training on Level 3, all three participants were also re-assessed on cognitive (i.e., SB5) and language (i.e., PLS4) assessments. On the SB5, participant 1 showed an increase in Age Related Potential (ARP) of four months from pre-post test, the second participant showed an increase of nine months, and participant 3 showed a decrease of one month. On the PLS4, all children demonstrated large increases in ARP from pre to post-test. The first participant showed an increase of 23 months and participants 2 and 3 of 13 months and 24 months respectively. Moreover, with the exception of one child (who demonstrated an increase of 15 months), these rises represented the largest increases from pre to post test in Study 5. As participants were not assessed following completion of Level 2 training, the improvements cannot be attributed to Level 3 alone. Nevertheless, these data support the idea that relational training more generally is beneficial in terms of intellectual potential.

These findings also provide further demonstration of the utility of MET for training REP responding in young children. In doing so, they extend the results from Study 8 by showing that the relatively complex pattern of relational responding involved in Level 3 of the NSD-REP can be taught even to a relatively young age group. More specifically, they indicate the potential for training complex relational responding in young age groups by breaking such responding down into simpler patterns as a precursor. Given the potential of relational training for improving intellectual potential and of the REP as means of providing it, this is useful to see.
Study 9 Data

Sessions

% of Correct Responses

Pt 9.1

Pt 9.2

Pt 9.3

Not Same/Not Diff (Test);
Same/Not Same (Train);
Diff/Not Diff (Train);
Not Same/Not Diff (Train).

Complex Relational Responding & Using the REP in Mid-Childhood
**Complex Relational Responding & Using the REP in Mid-Childhood**

*Figure 8.3. Study 9. Data for participants (Pts 1-3) in Study 9 MBD with NSD-REP Level 3 (not same/not different) responding (Phases 1,5 and 6) as the target and with multi-phase (2-4) intervention training. This shows accuracy in responding to training and testing over exposures in Phase 1 (baseline, Stimulus Set 1); Phase 2 (same/not same, multiple exemplar training, novel stimulus sets); Phase 3 (different/not different, multiple exemplar training, novel stimulus sets); Phase 4 (not same/not different, multiple exemplar training, novel stimulus sets); Phase 5 (return to baseline, Stimulus Set 1); and Phase 6 (follow-up, Stimulus Set 2).*

**General Discussion**

The current series of studies aimed to explore the use of the NSD-REP as a means of assessing and teaching children in middle childhood (age range 4-7) to respond to varying levels of contextual control over non-arbitrary same and different relations and featuring in particular negation control. To accomplish this, Study 7 assessed 23 children on the NSD-REP as well as on several measures of intellectual potential including the SB5 (an intelligence test), the RAI (a measure of relational framing), the DPRT-R (standardised reading measure) and DPMT-R (standardised maths measure) and showed correlations between the NSD-REP and other measures. Following this, Study 8 used the NSD-REP protocol to successfully train three of the children from Study 7 who had failed Level 3. Finally, Study 9 extended the findings of Study 8 by successfully training three further children at a much younger age (i.e., four years) in the same repertoire. Overall, the results further support the suggestion that the REP and more specifically the NSD-REP can be used to effectively assess and train contextually controlled relational responding in young children.

One of the outcomes of the work reported in this chapter was the finding of a relatively strong positive correlation between the NSD-REP and other standardised
measures of intellectual potential including a measure of cognitive ability / intelligence (i.e., the SB5) and one standardised school assessment, the DPMT-R (i.e., Drumcondra Math); this time the children involved had a mean age of approximately seven years, a number of years older than the mean age of the oldest group in any of the previous studies. This further supports the use of the NSD-REP as a measure of intellectual potential based on its ability to tap into relational responding as a key repertoire underlying intellectual performance and it extends the possibility of using the NSD-REP as a relatively simple and brief assessment for this purpose. Interestingly, the NSD-REP did not correlate with the RAI and the RAI did not correlate with any of the other, standardised assessments either. As discussed earlier, the participants found this protocol very difficult. Perhaps the children may have been too young and thus: (i) the level and/or the range of derived relational responding required may have been unsuitable; and/or (ii) the test may have required a duration of on-task behaviour that was relatively extended compared with the previous experience of this population, with the result that the level and range of scoring were depressed and as such there was much less likelihood of correlation with alternative measures. In explanation of the contrast between the RAI and NSD-REP, the latter tests for a much more restricted relational repertoire than the former and one that even relatively young children might be likely to have. Though with varying degrees of contextual control over demonstration of that repertoire in turn required to perform well at the upper levels, the NSD-REP is still a potentially testing measure for children with relatively higher relational skills. Additionally, this REP inspired protocol is briefer to explain and administer, which may also have helped. As suggested earlier, this might indicate that the NSD-REP would be a more
reliable indicator / predictor of intellectual potential for a younger age group than the RAI.

One interesting and potentially useful feature of the NSD-REP is that it taps into negation. Negation is an important aspect of a relational repertoire that children need to acquire. As discussed in the introduction, it has been found in several studies that there appears to be delayed comprehension of negative sentences compared to positive (or un-negated) sentences (e.g., Carpenter, 1971; Carpenter & Just, 1975). The results of the present program of research might be seen as showing similar results in a broad sense as the relatively older children involved in the current study performed well on Level 3 without any training whereas younger children in some of the previous studies failed Level 3 despite doing reasonably well on lower levels. As suggested in the introduction, perhaps younger children have received insufficient MET opportunities to come under appropriate negation-type control. It is also possible that their performance is a function of the additional complexity of the contextual control involved in tests such as Level 3 of the NSD-REP; they may be less likely to come under appropriate control due either to less experience of language/relational responding in general or of complexity per se. These are issues which further exploration using the NSD-REP or similar methods with various age groups might investigate.

In addition, a number of children attained the maximum score on the NSD-REP suggesting that the negation control involved in Level 3 was not particularly testing for them. In fact Level 3 of the NSD-REP as it stands should not have been particularly difficult for children reasonably experienced in responding under the control of cues for negation. This is because negation responding was required in all trials and thus, while certainly making the task involved more difficult than a task
without negation cues, the level of performance required was not as high as a task in which trials might vary in terms of the presence or absence of a cue for negation in addition to varying along the dimension of relation type (i.e., between same and different). Such a level was not created in the context of the present research program given the focus on younger children who would likely not have performed well on it. However, research on children’s performance in the context of this type of switching of contextual control is important. For example, one of the reasons described previously for researching the REP with children is because of the fact that this format has given rise to various protocols that have been used to assess and/or train potentially important examples/properties of relational responding in adults which we might be interested to also examine in children. One of those mentioned in the introduction for example was the IRAP, which is an example of the REP in which switching of contextual control is deliberately employed for various purposes including tapping into previous history of relational responding (e.g., Vahey et al., 2009) as well as assessment and training of cognitive performance (see e.g., O’Toole & Barnes-Holmes, 2009). Indeed in the introduction it was suggested that it was important to examine responding under control of negation cues because such cues were common in the IRAP. Switching contextual control of negation cues in addition to other cues across trials is also important in applications of the IRAP and thus to fully test capacity for IRAP performance, such switching should also be investigated in children. This is a recommendation for future research.

Apart from allowing comparison with alternative assessments and giving rise to theoretical issues with regard to development and training of particular repertoires, the outcome of the NSD-REP assessment in Study 7 also provided the basis for identifying individuals suitable for an experimental intervention. The ensuing
studies (8 and 9) allowed training of children on Level 3, which featured relatively more complex contextual control including negation. In Study 8, three participants who met criterion for NSD-REP Level 2 but not for Level 3 were randomly selected and trained to meet criterion for the latter employing a concurrent multiple-baseline design (MBD) across participants. All three participants in this study seemed to respond quite quickly to the intervention and thus it seemed particularly effective. This might also have been related to age of course as all three were seven year olds. Accordingly, in Study 9, a similar intervention was conducted with much younger children with a mean age of between 4 and 5 years. Once again, training successfully established Level 3 responding in all three children involved. These findings extend those from Study 8 by showing that the relatively complex pattern of relational responding involved in Level 3 of the NSD-REP can be taught even to a relatively young age group. In the case of this study, training needed to be broken down into smaller steps in order to establish the final repertoire. The fact that this resulted in success demonstrates the potential for training complex relational responding in young age groups by breaking such responding down in this way.

Taken together the findings from Studies 8 and 9 further support the efficacy of MET as a means of training Level 3 NSD-REP performance. In doing so, they extend the findings from Studies 2-4 and 6 in Chapters 6 and 7 respectively. They do so by providing further evidence of the utility of MET of relational responding. They also indicate that the effects of training are maintained several weeks later with new stimulus sets. Given the correlation with IQ and mathematical skill shown in Study 7, improving this repertoire is potentially important. Thus, the results of the training studies herein which illustrate that it can be successfully trained and maintained with totally novel stimulus sets, are encouraging.
Summary

In conclusion, the studies described in the current chapter provide additional evidence for the potential contribution of the NSD-REP for assessing and training relational responding in relatively young children. This is consistent with previous studies reported in the current thesis, but this work was focused on assessment and training of more complex relational responding that depended to an important extent on negation responding. An interesting finding to emerge was that while NSD-REP protocol significantly correlated with cognitive and school based standardised assessments for the younger children involved in these studies, the RAI did not. As discussed, this is understandable, as the levels of relational responding and on-task behaviour required by the RAI are quite a bit higher than those required by the NSD-REP and thus, this methodology may simply be more suitable as an assessment tool for this age range. Of course, the NSD-REP focuses on a relatively narrow range of relational responding and a narrower range than the RAI. However, what assessment and training studies such as those featured in this and previous chapters show is the potential that the REP has as an assessment and training tool with relatively young children. If results such as these can be achieved with an REP focused on simple non-arbitrary same and difference relations, then expansion to a wider field of relations and to arbitrary relations seems feasible and is certainly worthy of future research. Such work might help assessment with respect to aspects of relational repertoires already present (e.g., particular deficits, particular biases) as well as helping to establish wholly new repertoires (e.g., in respect of types of relations either entirely absent or at most weakly emergent; these might range from relatively simpler relations such as comparison or opposition to relatively more complex repertoires such as hierarchy or analogy).
Chapter 9

Study 10

Comparing the effects of derived relational training and computer coding on intellectual potential in school age children

* This chapter has been published: Hayes, J., & Stewart, I. (2016). Comparing the effects of derived relational training and computer coding on intellectual potential in school age children. *British Journal of Educational Psychology.*

doi:10.1111/bjep.12114
The overall aim of the current programme of research has been to advance the use of the REP with children, both neurotypical and developmentally delayed. Utilising the NSD-REP, the findings from Studies 1-9 indicate the utility of the REP generally, and of this multi-stage REP protocol in particular, as a means of measuring and training relational ability across a range of ages (i.e., 2.5-7 years). The final study in this thesis also investigated the use of an REP protocol with children, but it differed from the previous research in three ways: (i) it focused on a relatively older range of children; (ii) it used a more sophisticated version of the REP, one that includes a greater range of relations (both arbitrary and non-arbitrary); and (iii) it employed a group-based design. Prior to explicating Study 10, the rationale for this investigation is outlined in more detail. Following this, the research methods are described, the results are presented and then the findings and their implications are discussed.

**Rationale for Study 10**

In the last chapter (i.e., Study 7), the NSD-REP was correlated with intellectual measures and a more complex relational procedure, namely the Relational Abilities Index (RAI). The results revealed a comparatively strong positive correlation between the NSD-REP and standardised school and IQ assessments. Interestingly, however, the RAI did not relate to any other measure, not even the NSD-REP itself. Though these findings are speculative, they seem to suggest that the NSD-REP is a more useful tool for assessing and possibly training younger children, while the RAI is better suited for use with older children (Cassidy, Roche, & Hayes, 2011).

In order to advance the REP work conducted thus far, a different type of REP methodology was utilised in the current study with children aged 10-12 years. The
protocol in question, is SMART (Strengthening of Mental Abilities using Relational Training), an online program for assessing and training derived relational responding (DRR). In previous research, Cassidy and colleagues (2011) found that an early prototype of the SMART correlated with intelligence measures in children aged 8-12 years. More importantly, they highlighted that multiple exemplar training (MET) using this methodology demonstrated significant improvements in IQ scores from pre to post-test (for a detailed description of this study please refer to p. 61).

While the data from Cassidy et al. (2011) are highly promising at both a theoretical and a practical level in terms of the apparent effect of DRR on intellectual potential, this study had a number of limitations. These encompassed a very small sample size, the absence of an active control group and restriction in terms of the measurement of academically, as opposed to purely intellectually relevant outcomes. As such, additional exploration of the potential of derived relations training to boost intellectual potential is needed. The current study attempted to overcome the aforementioned weaknesses by exploring the effect of derived relations training under more controlled conditions.

First, this study featured a substantially larger sample of 28 children, with 14 in the relational frame training group and fourteen in the control group. Second, it included an active control condition who received exposure to an intellectually demanding training regime (in “Scratch” computer coding) for the same amount of time as the experimental condition received exposure to derived relations training. The latter was provided using the SMART, whose core content is based on the multiple stimulus relations training used in Cassidy et al. (2011). Scratch, meantime, was developed by researchers at MIT and previous research has shown positive effects of Scratch exposure on problem-solving abilities and logical
reasoning skills in children (Lai & Yang, 2011; Shin & Park, 2014). In order to ensure equality of exposure across the two conditions, the same amount of training time (i.e., 29 hours in total) was allowed for both. Third, as regards outcomes, participants were measured before and after training on IQ tests similar to those used in the Cassidy et al. study but importantly they were also measured along academically relevant dimensions of mathematical and linguistic (including reading, writing and oral) ability.

Hence, the final study reported in this thesis focused on the assessment and training of REP performance in pre-teens. It sought to extend the REP research presented in the thesis thus far, by training and testing relational responding in an older group of children. In addition, it aimed to extend the Cassidy et al. (2011) study by using an REP protocol under more controlled conditions and investigating whether this training had an effect on IQ and standardised school assessments.

Method

Participants and setting

Twenty-eight children (13 girls) aged 10–12 years \((M = 10.34)\) participated. The participants were all members of the same class in a primary school in Co. Limerick, Ireland. Permission to send consent forms to caregivers was obtained from the school principal and class teacher. Participants were included if consent forms were signed and returned and if the child assented. Everyone in the class participated on this basis.

All participants were Caucasian. The majority \((n=24; 86\%)\) were Irish nationals, with Polish \((n=3; 11\%)\) and Russian \((n=1; 3\%)\) nationals making up the remainder. Six students \((22\%)\) were in receipt of special remedial education services in reading and/or math (typically just over four hours of supplementary support per
week) and had Individualized Education Plans (IEPs). None of the students in the class had visual, auditory, or physical disabilities that impeded their progress in the general education curriculum. During the course of the study, 14 children (7 female, $M = 10.34$ years, $SD = 0.45$) were randomly assigned to the experimental (SMART) condition and 14 (6 female, $M = 10.34$ years, $SD = 0.44$) to the control (Scratch) condition. Based on this, three of those in receipt of remedial educational services (P2, P5 and P23) joined the SMART group, while the remainder (P10, P13, P19) joined the Scratch group. One of those receiving extra support in the SMART group (P23) had a formal diagnosis of Dyslexia. A second (P5) had what was recorded as a “difficult home background”. P2 and P23 were in receipt of supplementary support in math while P5 received assistance in reading as well as art therapy. Of the three students in the Scratch group receiving supplementary help, two (P10 and P13) received assistance in reading while the third (P19) received support in math. None of the latter three students had any formal diagnoses at the time of this study.

Materials & Apparatus

All participants were tested on several assessments before and after the respective interventions. The first assessment was the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). This is an individually administered IQ test with four sub-tests, two (“Vocabulary” and “Similarities”) contributing to verbal IQ (VIQ) and two (“Block Design” and “Matrix reasoning”) contributing to performance IQ (PIQ). Scores on all four sub-tests contribute to a FSIQ score. The full battery requires 30 min to administer and has an adequate reliability. Mean internal consistency reliabilities for 6- to 16-year-olds are .96 and .93, respectively, for the FSIQ-4 and FSIQ-2 and .93 and .94, respectively, for the VIQ and PIQ.
Stability coefficients for the same age range are .93 for the FSIQ-4, .85 for the FISQ-2, .92 for the VIQ, and .88 for the PIQ (Wechsler, 1999).

Participants were tested on three sub-tests of the WIAT-II (Wechsler Individual Achievement Test Second UK Edition; Wechsler, 2005). This is an individually administered achievement test used for diagnostic purposes and includes nine sub-tests of various academic skills. The three sub-tests used here (reading, spelling, and numerical operations) were chosen because, while correlated with general intellectual ability, they are also directly relevant to academic performance. Average subtest reliabilities for each of these sub-tests are above .90. Test–retest stability for these sub-tests also exceeded .90.

Two sub-tests of the Wechsler Intelligence Scale for Children (4th Edition) (WISC-IV UK; Wechsler, 2004) were also assessed. This is an individually administered clinical instrument for assessing children’s intellectual ability and comprises 13 sub-tests. The two sub-tests used here (letter/number sequencing and digit span) were chosen because they tap into “working memory”. Reliability coefficients for digit span and letter/number sequencing are .87 and .90, respectively. Test–retest stability coefficients for both sub-tests are .83 (Williams, Weiss, & Rolfhus, 2003).

The above three IQ assessments were administered to all members of the class on an individual basis by the main experimenter immediately before and immediately after the intervention. Two additional measures were nationally standardized measures that were group-administered by teaching staff at the school. The Drumcondra Primary Reading Test-Revised (DPRT-R) and the Drumcondra Primary Math Test-Revised (DPMT-R) are group-administered, standardized tests.
of achievement in reading and mathematical ability, designed for pupils in Irish primary schools. For a description of both assessments, please see pp.162-163.

Participants were also tested before and after on the relational abilities index (RAI). This was delivered online using the same website as the SMART training intervention. Again, this assessment has been described in detail elsewhere in the thesis (please see p. 164). This program recorded an RAI score based on the total number of correct responses made. Reliability was calculated using data from the Scratch (control) group, who received no relational training between pre- and post-testing. The single measure intraclass correlation index was .932 with a 95% confidence interval from .806 to .978, F(13, 13) = 27.368, p < .001.

Both interventions (i.e., Scratch and DRR/SMART) were delivered via modern computer equipment. The DRR/SMART was delivered via Apple iPad tablets. The Scratch training was delivered via Dell Latitude laptop computers.

**Design**

The experiment used a 2 x 2 mixed-model design to evaluate and compare two conditions (experimental [SMART] and control [Scratch]) on several dependent measures (WASI, RAI, Drumcondra [Reading, Maths], WIAT-II, and WISC-IV) both before and after they received their respective intervention. After pre-intervention assessments, participants were randomly assigned to one of the two conditions and they were assessed on the same measures once more following the intervention. Data were analysed in SPSS-20 (Statistical Package for the Social Sciences, Version 20; IBM Corporation, Armonk, NY, USA) using a two-way mixed-design multivariate analysis of variance (MANOVA).
Procedure

**Baseline Assessment.** All pre-intervention assessments except for the Drumcondra tests were administered in a small room adjacent to the general education classroom in the participants’ school. All 28 participants were assessed on an individual basis in this room on several measures described in the previous section including: (1) the RAI; (2) four WASI-II sub-tests; (3) two WISC-IV sub-tests; and (4) three WIAT-II sub-tests. These were administered in the same order for each child, and testing sessions took approximately 80 minutes in total. After all children had been assessed, they were randomly assigned to one of the two conditions with equal numbers (n = 14) in each.

**Intervention: Strengthening mental abilities with relational training (SMART).** Training for the SMART (derived relations training) condition occurred in the science laboratory of the participants’ school. It was conducted on a group basis, with all participants involved. Participants were exposed to the online SMART protocol that trained responding to the derived relations of “Same/Opposite” and “More/Less”. For training purposes, they were brought from their classroom to the science laboratory for 1-hour instalments, typically twice per week. Total training time involved 29 hours worth of sessions over two academic semesters. The length of training for this group was yoked to that provided for the Scratch condition, who were provided with this length of training as a precursor to submission of their coding projects in the context of a national “coding” competition.

In order to ensure that all SMART participants received 29 hours of training, a daily log was taken by the researcher. If participants were absent during training hours, the researcher either set up another training session in school or alternatively, participants were given permission to log into their accounts from...
home and complete the session they had missed (again this was monitored by the researcher). Please refer to Table 9.1 below for a review of weekly sessions, number of sessions missed by each participant, additional training sessions at school and home and total number of levels completed in SMART.

For training purposes, all participants were seated at a table with an individual iPad. In the first session, they were taught how to access the online protocol. For subsequent sessions, website address and access instructions were written on the blackboard and the experimenter asked everyone to log in and ensured that all could access the training. From session 3, all participants were given three minutes to log in.

The online SMART training protocol involves seventy levels that assess and train DRR. Thirty seven levels train “Same/ Opposite” while thirty three train “More/Less”. For each type of relation, levels increase progressively in difficulty over the course of the levels for that relation. Each level includes a training and a test phase. During training, the learner must respond correctly to 16 consecutive exemplars of that level, within a 30 seconds time limit. Tasks continue ad infinitum until this criterion is reached. Auditory-visual feedback (“Correct” or “Wrong”) is provided. In testing, the learner must respond correctly to a single finite block of 16 consecutive exemplars without feedback. If they pass, they move onto the next level. If they fail, they have to repeat both the training and test phases for that level.
Table 9.1

*Number of weekly sessions for participants, number of sessions missed, additional sessions in school and home and Level attained in the SMART condition*

<table>
<thead>
<tr>
<th></th>
<th>Pt 1</th>
<th>Pt 2</th>
<th>Pt 3</th>
<th>Pt 4</th>
<th>Pt 5</th>
<th>Pt 6</th>
<th>Pt 7</th>
<th>Pt 8</th>
<th>Pt 9</th>
<th>Pt 10</th>
<th>Pt 11</th>
<th>Pt 12</th>
<th>Pt 13</th>
<th>Pt 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wk 1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Wk 2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Wk 3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Wk 4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Wk 5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Wk 6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Wk 7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Wk 8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Wk 9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Wk 10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
### Derived Relations Versus Coding for IQ

<table>
<thead>
<tr>
<th></th>
<th>Wk 11</th>
<th>Wk 12</th>
<th>Wk 13</th>
<th>Wk 14</th>
<th>Wk 15</th>
<th>Wk 16</th>
<th>Wk 17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sessions Missed</td>
<td>2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td>
<td>2 2 2 0 2 1 2 2 2 0 2 1 0 2</td>
<td>2 2 2 2 2 1 1 2 2 1 2 1 1 1 1 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ School Sessions</td>
<td>1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1</td>
<td>2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Home Sessions</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Hours</td>
<td>29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29</td>
<td>29 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level Attained</td>
<td>61 36 (st) 62 62 62 29 31 (st) 62 34 25 66 26 (st) 27 30 (st)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Pt = Participant; Wk = Week; + School/Home Sessions = The number of additional school or home sessions provided due to absence; Level Attained = The level participants attained by the end of training; st = The second time around
**Intervention: Scratch.** Training for the Scratch group took place in the general education classroom in the children’s school. Training was provided by the class teacher and organized into a series of lessons whose ultimate goal was to guide participants in the use of Scratch computer coding to create a computer game. As for the other group, participants received a total of 29-hour training spread across two academic semesters in twice-weekly 1-hour installments.

In Lesson 1, participants were introduced to the Scratch interface and were facilitated in creating their own “Sprite” (character). They were also exposed to the basic building blocks of coding that would enable them to write code that would move and alter their Sprite (Figure 9.1). Causing horizontal or vertical movements required an understanding of coordinate geometry and integers; up/down movements required changes to the y-axis, while left/right required changes to the x-axis; movement right required an increase in x, while movement left required a decrease in x; similarly, movement up required an increase in y, while movement down required a decrease in y. Coding was written so that particular keys would be assigned particular movements. These and other related concepts were trained and reviewed each week.

Over the first 14 hours of the course, learners were taught to write code to develop various character costumes; play music; and create a quiz, respectively. In the remaining 15 hours, they had to develop their own game. This task, the most challenging part, required the integration of previously learned skills. In this latter phase, the teacher guided and encouraged the participants in their activities.
Figure 9.1. Example of the Scratch computer coding interface.

**Post-Assessment.** The post-test phase was identical to baseline.

**Results**

**Main Analysis: Descriptive**

Descriptive data for RAI, WASI, WIAT, WISC and Drumcondra measures for the two conditions at both pre- and post-intervention are provided in Table 9.2 and represented in Figure 9.2. Raw rather than standardised scores are provided in the case of all measures so that ability as opposed to ability in relation to age is assessed. As may be seen, there are improvements in the mean performance of both conditions in the case of many of the measures. However, the experimental (SMART/DRR) condition improves on more measures than the control (Scratch) condition and also shows noticeably larger improvements than the latter in a number of these cases.

The three students in the SMART group who were receiving supplementary help performed comparatively well. P23 (diagnosed with dyslexia) showed improvements across all psychometric measures and demonstrated the third highest
improvement in the class on the WIAT reading subtest. P5 (‘difficult background’) started with the lowest WASI-measured IQ (75) and improved by nine points, which was larger than the average for the SMART group as a whole. Finally, P2 improved on all IQ measures and achieved the highest improvement in the class on the WIAT reading subtest. As a gauge of the performance of this subset of the SMART group, we compared their improvement scores with the average improvement of the whole of the “Scratch” group and found that they outperformed them on every index except for the Drumcondra Math and English tests.
Table 9.2

Mean performance for the experimental and control group on each dependent measure (raw scores) at pre and post-test

<table>
<thead>
<tr>
<th>DV</th>
<th>Experimental (SMART)</th>
<th>Control (Scratch)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-Test M (SD)</td>
<td>Post-Test M (SD)</td>
</tr>
<tr>
<td>RAI</td>
<td>30.9 (6.9)</td>
<td>40.6 (8.6)</td>
</tr>
<tr>
<td>WASI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>34.3 (6.4)</td>
<td>39.5 (7.5)</td>
</tr>
<tr>
<td>Block Design</td>
<td>28.4 (14.1)</td>
<td>39 (16.8)</td>
</tr>
<tr>
<td>Similarities</td>
<td>24.0 (4.6)</td>
<td>27.0 (3.7)</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>21.6 (4.3)</td>
<td>23.7 (3.1)</td>
</tr>
<tr>
<td>WIAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling</td>
<td>30.6 (4.2)</td>
<td>37.0 (4.0)</td>
</tr>
<tr>
<td>Reading</td>
<td>102.1 (6.8)</td>
<td>111.2 (4.7)</td>
</tr>
<tr>
<td>Num. Operations</td>
<td>23.8 (2.2)</td>
<td>29.6 (4.3)</td>
</tr>
<tr>
<td>WISC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span.</td>
<td>13.1 (2.7)</td>
<td>16.9 (3.0)</td>
</tr>
<tr>
<td>Letter/Number Seq.</td>
<td>13.0 (4.2)</td>
<td>18.9 (1.0)</td>
</tr>
<tr>
<td>DRUMCONDRA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>50 (15.2)</td>
<td>48.2 (15)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>48.5 (16.1)</td>
<td>49.5 (16.4)</td>
</tr>
</tbody>
</table>
Figure 9.2. Graphs comparing mean performance for the Experimental and Control groups at both pre- and post-testing on each of the dependent measures.
Main Analysis: Inferential

A two-way mixed design multivariate analysis of variance (MANOVA) was used to compare outcomes for the two conditions. The results indicated a within subjects effect of time, $F(12,15) = 25.4$, $p < .001$, $\eta^2 = 0.95$, and an interaction of time*condition, $F(12,15) = 13.9$, $p < .001$, $\eta^2 = 0.92$.

Univariate follow up tests examining for the effect of time alone found significance for several variables including relational ability, $F(1,306.4) = 25.96$, $p < .001$, $\eta^2 = 0.5$; WASI Vocabulary, $F(1,224) = 27.69$, $p < .001$, $\eta^2 = 0.516$; WASI Block design, $F(1,721.5) = 24.93$, $p = .001$, $\eta^2 = 0.49$; WASI Similarities, $F(1,111.5) = 33.68$, $p = .001$, $\eta^2 = 0.56$; WASI Matrix reasoning, $F(1,80.2) = 10.67$, $p = .003$, $\eta^2 = 0.29$; WIAT spelling, $F(1,224) = 54.32$, $p < .001$, $\eta^2 = 0.68$; WIAT reading, $F(1,516) = 87.17$, $p < .001$, $\eta^2 = 0.77$; WIAT numerical operations, $F(1,157.8) = 34.87$, $p < .001$, $\eta^2 = 0.57$; WISC digit span, $F(1,66.5) = 53.9$, $p < .001$, $\eta^2 = 0.68$; and WISC letter number sequencing, $F(1,114.3) = 26.1$, $p < .001$, $\eta^2 = 0.50$.

Univariate follow up tests examining for time*condition interaction effects found significance for several variables including relational ability, $F(1,365.2) = 30.94$, $p < .001$, $\eta^2 = 0.54$; WASI block design, $F(1,385.9) = 13.34$, $p = .001$, $\eta^2 = 0.34$; WIAT spelling, $F(1,77.8) = 18.86$, $p < .001$, $\eta^2 = 0.42$; WIAT reading, $F(1,126) = 12.28$, $p < .001$, $\eta^2 = 0.45$; WIAT numerical operations, $F(1,82.6) = 18.25$, $p < .001$, $\eta^2 = 0.41$; WISC digit span, $F(1,33) = 26.80$, $p < .001$, $\eta^2 = 0.51$; WISC letter number sequencing, $F(1,126) = 28.81$, $p < .001$, $\eta^2 = 0.53$; and Drumcondra mathematics, $F(1,52) = 4.75$, $p = .038$, $\eta^2 = 0.16$. 
Additional independent t-tests were conducted to compare the improvement from pre- to post-test of the two groups on each of the variables implicated in the time*condition interaction. These showed significant differences between the two for improvement in relational ability, \( t(26) = 5.562, p < .001 \); in WASI block design, \( t(26) = 3.652, p = .001 \); in WIAT spelling, \( t(26) = 4.343, p < .001 \); in WIAT reading, \( t(26) = 4.613, p < .001 \); in WIAT numerical operations, \( t(26) = 4.272, p < .001 \); in WISC digit span, \( t(26) = 5.177, p < .001 \); in WISC letter number sequencing, \( t(26) = 5.367, p < .001 \); and in Drumcondra mathematics, \( t(26) = 2.180, p = .038 \). As is confirmed by Table 9.2 and Figure 9.2, in all cases these differences favoured the Experimental (SMART/DRR) condition.

**IQ (WASI) Analysis: Descriptive & Inferential**

Full scale WASI IQ scores for the two conditions for the two conditions at both pre- and post-intervention are shown in Figure 9.3. As may be seen, there is a slight (non-significant) difference between the experimental (98.9) and control (97) conditions (i.e., 1.9 points) before training, but the post-training difference between the conditions (favouring the SMART condition) is considerably larger (i.e., 107.3 and 98.1 respectively or a difference of 9.2) and significant \( (p < .05) \). As may also be seen, there are improvements in the mean performance of both conditions with the experimental (SMART) condition improving more than the control (Scratch) condition. A two way mixed analysis of variance (ANOVA) was used to compare the two conditions. The results indicated a within subjects effect of time, \( F(1,26) = 34.6, p < .001, \eta^2 = 0.57 \), and an interaction of time*condition, \( F(1,26) = 18.5, p < .001, \eta^2 = 0.41 \).
Correlations between derived relational ability as measured by the RAI and ability on other assessments were measured across both groups at pre-intervention (see Table 9.3). As may be seen, RAI scores correlated significantly with a number of measures including WASI block design ($r = .45$) and full scale IQ ($r = .44$), WIAT spelling ($r = .52$), reading ($r = .49$) and numerical operations ($r = .4$), WISC letter number sequencing ($r = .41$) and Drumcondra English ($r = .59$) and Mathematics ($r = .69$).
Study 10 sought to extend the research presented thus far by examining the potential of a sophisticated REP protocol, namely SMART, as a means of training and testing an older cohort of children (i.e., aged 10-12 years). It also served as a follow-up to a previous published study which found that: (i) children’s ability to derive arbitrary relations, using an REP type protocol, correlated with measured IQ; and (ii) that training DRR produced substantial and unprecedented IQ rises.

Table 9.3.

Correlation (Pearson’s r) of RAI & Other Measures at Pre-intervention

<table>
<thead>
<tr>
<th>DV</th>
<th>r</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WASI Vocabulary</td>
<td>.34</td>
<td>.074</td>
</tr>
<tr>
<td>WASI Block Design</td>
<td>.45</td>
<td>.017*</td>
</tr>
<tr>
<td>WASI Similarities</td>
<td>.28</td>
<td>.156</td>
</tr>
<tr>
<td>WASI Matrix Reasoning</td>
<td>.29</td>
<td>.141</td>
</tr>
<tr>
<td>WASI IQ (Full)</td>
<td>.44</td>
<td>.018*</td>
</tr>
<tr>
<td>WIAT Spelling</td>
<td>.52</td>
<td>.005**</td>
</tr>
<tr>
<td>WIAT Reading</td>
<td>.49</td>
<td>.008**</td>
</tr>
<tr>
<td>WIAT Numerical Ops</td>
<td>.40</td>
<td>.033*</td>
</tr>
<tr>
<td>WISC Digit Span</td>
<td>.23</td>
<td>.239</td>
</tr>
<tr>
<td>WISC Letter/No. Seq.</td>
<td>.41</td>
<td>.031*</td>
</tr>
<tr>
<td>Drumcondra English</td>
<td>.59</td>
<td>.001**</td>
</tr>
<tr>
<td>Drumcondra Math</td>
<td>.69</td>
<td>.0001***</td>
</tr>
</tbody>
</table>

Discussion

Study 10 sought to extend the research presented thus far by examining the potential of a sophisticated REP protocol, namely SMART, as a means of training and testing an older cohort of children (i.e., aged 10-12 years). It also served as a follow-up to a previous published study which found that: (i) children’s ability to derive arbitrary relations, using an REP type protocol, correlated with measured IQ; and (ii) that training DRR produced substantial and unprecedented IQ rises.
The findings indicate that, similar to the research presented in Chapters 6 and 8, and previous published research (i.e., Cassidy et al., 2011), training in DRR positively affected the intellectual potential of 10-12 year old children. All of the children who were trained using the SMART online system showed significant improvements across a variety of measures of intellectual ability and academic attainment when compared with the control condition (i.e., Scratch). This included several children in receipt of supplementary remedial training, including one who had the lowest IQ in the class (75) and improved by nine points and a second who had Dyslexia and showed the third best improvement in the class on the WIAT reading sub-test. In addition, by showing once again the influence of training DRR on intellectual potential, Study 10 extends Cassidy et al. (2011) in several respects.

First, a much larger sample of participants (n=28) was employed, which adds to the confidence concerning the reliability of the results. Second, an active control condition was utilised. In the “Scratch” group, participants were engaged in computer-based intellectual training throughout the period in which the experimental condition received training in DRR. Additionally, as part of the attempt to control for extraneous variables, both conditions received precisely the same quantity of training (i.e., 29). Third, all participants were tested before and after the intervention stage, not only on several different IQ type sub-tests, but also on academically relevant dimensions of mathematical and linguistic (including reading, writing and oral) ability.

When taken together, the results demonstrated that training in DRR using the SMART protocol can foster broadly assessed intellectual skills. Examination of change from pre- to post-intervention, without consideration of condition, showed improvement for relational ability, for full scale WASI IQ, for each of the WASI
sub-tests and for each of the sub-tests in the other measures. Examination of time by condition interaction effects showed improvements for the SMART condition over the Scratch condition for relational ability, for full scale WASI IQ, for most of the IQ sub-tests and also for the Drumcondra Mathematics performance while follow-up tests confirmed these patterns where they emanated from multivariate analysis.

The difference between the SMART and Scratch conditions in full scale IQ, as well as across several of the intellectual sub-tests, provides further evidence of the promise of training DRR for improving intellectual potential. The additional tests used in this study as compared with Cassidy et al. (2011) also allowed for more detailed evidence in this respect. As previously noted, this study used several different IQ type sub-tests as well as academically relevant dimensions of mathematical and linguistic (including both writing and reading) ability. The results showed greater improvement for SMART not only in WASI IQ but in several of the academically focused IQ sub-tests (i.e., WIAT spelling, reading and numerical operations, WISC letter-number sequencing) as well as in the Drumcondra standardised test of achievement in mathematics. It is notable that improvement was seen across all of the tests relevant to mathematics ability, suggesting that DRR training may be particularly important for mathematical skill, which would perhaps not be surprising given the extent of arbitrary relations involved in mathematics. However, the fact that improvement extended to several other sub-tests relying on other types of academic skill (e.g., WIAT spelling and reading) also supports the idea that DRR is important for academic skills in general.

As a further index of the core relationship between DRR and intellectual performance, correlations were also found between relational ability and performance on each of the other assessments at pre-test across all participants. The
pattern of correlations was interesting and mirrored to some extent the pattern of differences found between the groups. For example, examining the correlations between relational ability and performance on the WASI sub-tests, “Block Design” had a much higher correlation than any of the others, while it was also the only subtest on which the SMART condition showed a significant difference from the Scratch condition. This adds to evidence that certain types of assessments are more closely associated with DRR ability than others.

Apart from IQ sub-tests assessing relatively academic skills, this study also employed a number of IQ sub-tests designed to gauge working memory capacity (i.e., the two WISC sub-tests of digit span and letter number sequencing) to check whether either intervention might affect this ability. An argument had been raised in Cassidy et al. (2011) that the key effect of training DRR was improvement in DRR skill rather than improvement in working memory capacity, and that the former rather than the latter was the major driver of improvement in IQ. However research is needed to see whether and to what extent working memory capacity might actually be improved by DRR training. On one hand it might be expected that DRR training would have at least some beneficial effect in this respect given that many of the tasks involved the sequential relation of a number of different elements. On the other hand, in each task all of the elements to be related were visible so that recollection of elements themselves was not necessary which would reduce burden on working memory. As it now seems, there was indeed some improvement in working memory based on SMART training and this might indeed mediate at least some of the improvement in IQ and other intellectual skills seen for this condition. Conversely, such a finding does not rule out the likelihood that at least some of the improvement
and possibility the majority, is due to improved ability to do DRR independent of working memory.

These results appear weaker in some respects than those of Cassidy et al. For example, in Cassidy et al., the average rise in (WISC) IQ from baseline after relational training was more than 22 points, whereas in the current study, the average rise in (WASI) IQ was just over eight points. The latter, while still being a very impressive jump in IQ, was quite a bit lower than shown in Cassidy et al. It seems likely that this is at least partly due to differences in the length of training and hence due to the absolute quantity of training received as well as the magnitude of the effect on other learning that took place concurrently. The relational training in the Cassidy et al. study was conducted twice weekly for substantial periods of time (not counting holiday periods and allowing for episodes of sickness etc.) over the course of an approximately two year interval and thus very likely involved over 200 hours of training in total (as a relatively conservative estimate). In contrast, the relational training provided in the current study was conducted once a week over the course of an approximately five month time interval (i.e., from approximately the start of November to approximately the end of March) and totalled 29 hours in total. In addition, the effect of an intervention such as DRR training is not simply a function of the number of hours of actual training themselves but also the effect of that training on other types of learning (in school and elsewhere) that are going on concurrently. Hence, the absolute difference in the number of hours indicates a potentially much larger difference based on this additional effect. This is likely at least a large part of the reason why, despite the fact that the SMART intervention is likely to be more efficient as a training tool than the intervention used by Cassidy et al. (2011), the effect sizes seen here are substantially smaller than in the latter.
A number of possible avenues for further investigation were also highlighted based on the results of the current study. First, it would be advantageous for future research to utilise larger randomized trials. Despite improvement over previous work, this study was still relatively small (n = 28) due to the fact that this was a sample of convenience based on a typical primary school class. Second, it will be important to improve the reinforcing value of the tasks in the SMART online programme. Even with the introduction of customised reinforcement strategies, some participants repeatedly failed particular levels, resulting in decreased motivation/performance. Third, it would be beneficial if a wider range of DRR skills could be targeted as part of this REP protocol including temporal (before–after), which has been explicitly linked with intellectual performance (e.g., O’Hora et al., 2008), as well as hierarchical, deictic (perspective-taking), and analogical relations. Fourth, follow-up tests would be beneficial to demonstrate real and lasting change in intellectual ability.

Together, the data reported in the current chapter provides additional evidence for the potential of the REP in assessing and training relational responding, this time with an older cohort of children.

**Summary**

In sum, the results of Study 10 extend the research presented in the literature and in the earlier chapters of this thesis by providing further empirical support of the importance of DRR for human linguistic and cognitive performance. These findings also highlight the potential for SMART in future investigations that examine the relationship between DRR and intelligence. Thus, although needing further replication and extension with yet better controls, the contribution of these results are arguably important both theoretically and practically.
Chapter 10

General Discussion
The core focus of the research presented in this thesis was to advance the use of the Relational Evaluation Procedure (REP), an innovative methodology designed to train and test for derived relations, with children across a broad spectrum of ages, with and without developmental delay. Over the course of a series of investigations, the utility of the REP in this respect has been explored, using controlled research designs. Studies 1-9 used a particular version of the REP, deliberately developed as a means of testing childrens’ ability to respond on an REP format, namely the NSD-REP (Non-arbitrary Same and Different Relational Evaluation Procedure). Conversely, the final study (i.e., Study 10) employed SMART (Strengthening Mental Abilities Using Relational Training), an online protocol that utilises an REP format to assess and train both non-arbitrary and arbitrarily applicable relational responding.

In this concluding chapter, a brief recap of the research area will be provided, in addition to the rationale and principle aims of the current research programme. Subsequently, the utility of the REP as a method for assessing and training relational framing and its potential to do so in children will be described. This will be followed by a summary of the thesis findings and how they relate to previous empirical investigations within the behaviour analytic literature. Next, the theoretical issues and implications of this research will be outlined, in addition to possible directions for future work. Finally, the chapter will finish with a brief concluding section.

Overview of Research Area and Aims of the Thesis

As detailed in the introductory chapters, the REP protocol is derived from Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), a contextual behaviour analytic approach to language and cognition. RFT explains
language and complex cognitive phenomena in terms of the operant of derived relational responding (DRR). For many years, the predominant method of studying this phenomenon within behaviour analysis has been the matching-to-sample (MTS) procedure. Yet, MTS as a methodology for studying derived stimulus relations is not conducive to the most effective progress of RFT research. It was for this reason that alternate methodologies were developed and arguably one of the more prominent and influential of these is the REP.

The REP facilitates exploration of relational responding by requiring participants to confirm or deny the applicability of particular stimulus relations to other sets of stimulus relations (Stewart et al., 2004). From an RFT perspective, this procedure has a number of advantages when compared to standard MTS methodologies. These include: (i) that it focuses on “relating” responses as opposed to “picking” responses and thereby lends itself to an analysis in terms of relations rather than classes; (ii) it can readily accommodate many different varieties of stimulus relations, in contrast to the MTS, which is best suited for studies of equivalence; and (iii) the patterns of relational responding that can be generated are not constrained by the prior training and testing of a specific set of derived relations, as is evident in MTS, and thus it lends itself to studying generativity, a key feature of language (Hayes & Barnes, 1997).

Not surprisingly, therefore, versions of the REP have been used to model and investigate a variety of different relations including temporal (“Before” and “After”) (e.g., O’Hora, Barnes-Holmes, Roche, & Smeets, 2004), coordinate and distinction (“Same” and “Different”) and analogical (Stewart et al., 2004) relations. In addition, the REP has evolved into, or at least inspired, several newer forms of protocol (e.g., the SMART; the Implicit Relational Assessment Procedure [IRAP]; and the
Relational Completion Procedure [RCP]), that are useful for research into DRR and hence the full range of complex human language and cognition (e.g., Cassidy, Roche & Hayes, 2011; Dymond & Whelan, 2010; Roddy, Stewart & Barnes-Holmes, 2010; Ronspies, Schmidt, Melnikova, Krumova, Zolfagari, & Banseet, 2015). Studies such as those cited have used the REP or its variants to provide additional evidence of the efficacy of RFT as an empirically based approach to human language and cognition and of the REP as an effective methodology for pursuing basic and applied RFT research.

In one respect, however, the potential of the REP has arguably not been adequately explored yet. This relates directly to REP research with children. Hence, the work presented in this thesis aimed to address this deficit by investigating the use of the REP as a means of assessing and training relational responding in children, including neurotypical children across a variety of age ranges, as well as children with Autism Spectrum Disorder (ASD). At this point, a summary on the work itself, including each study and its findings will be provided.

**Summary of Thesis Findings**

As described above, all studies, excluding one in the current research programme, used the NSD-REP, a version of the REP deliberately developed as a means of testing children’s ability to respond on this format. To quickly recap, the NSD-REP involves three levels. For Level 1, participants are required to evaluate non-arbitrary same and different relations by pressing either “Same” or “Different” in response to a pair of identical or non-identical pictures. For Level 2 pressing either “Yes” or “No” in response to a pair of identical or non-identical pictures presented together with an auditory recording of the word “Same” or “Different” is required. Finally, Level 3 is similar to Level 2 except control by the contextual cue
“Not” is required; for instance, a typical trial may involve two identical pictures presented onscreen with the discriminative stimulus “Are these not the same/different”, after which the participant is expected to press “Yes” or “No”.

Studies 1-5

Studies 1 to 5, presented in Chapter 6, examined the utility of the NSD-REP in young typically developing preschool children, ranging from 2.5 to 4 years. This was achieved by: (i) correlating scores on the NSD-REP protocol with performance on standardised, direct measures of language (i.e., the Preschool Language Scale 4th Edition [PLS4]; Zimmermann, Steiner & Pond, 2002) and cognition (i.e., the Stanford Binet 5th Edition [SB5]) (Study 1); (ii) training young children to respond accurately with respect to the relational evaluation of non-arbitrary same and different relations (Studies 2-4); and (iii) analysing the data from Studies 2-4 as a means of providing a preliminary gauge of the effects of NSD-REP training on the intellectual (linguistic and cognitive) potential (Study 5).

The results of Study 1 highlighted that children’s performance correlated (in a relatively large sample of 26 individuals) with that on standardised language and cognitive ability assessments. This, apart from suggesting the importance of relational responding for intellectual potential, enabled approximate categorization of capacity by age, as well as identification of individuals suitable for training in subsequent studies.

In the three studies that followed (i.e., Studies 2-4), children of varying ages and repertoires (n=12), derived from the larger sample, were trained using controlled designs, to perform on Levels 1 and 2 of the NSD-REP. Using multiple baseline designs (MBDs), the results of these investigations demonstrated that REP type relational repertoires, ranging from relatively simple to relatively complex, could be
effectively taught to young children. Though the younger children required more training sessions and the use of adjusted training procedures when compared to their older counterparts, training was successful for them also. Ultimately, the results of these studies illustrate that neurotypical children can be taught to respond to an REP interface at a very young age.

In Study 5, the participants that had received training across Studies 2-4 were re-assessed using the standardised measures from pre-testing in order to gauge whether training had a beneficial effect on their intellectual potential. Although tentative, there was at least some suggestive evidence of an increase in age equivalent scores on both the PLS4 and the SB5 from pre- to post-test, though well-controlled follow-up work is needed to be able to draw firmer conclusions. Taken together, however, the data presented in Chapter 6 adds to the growing empirical base that relational abilities are central to language and cognition and (more tentatively in the case of these particular training data) that training them can boost intellectual potential (e.g., Barnes, McCullagh, & Keenan, 1990; Cassidy, Roche, & Hayes, 2011; O’Hora, Pelaez, & Barnes-Holmes, 2005).

Study 6

Given the relative efficacy of the NSD-REP as a methodology for training relational skills in young typically developing children, one domain for extension of this work was to examine the efficacy of the NSD-REP in children with Autism Spectrum Disorder (ASD) and this served as the focus of Study 6. When this work commenced, there was only one published study, known to the author, examining the efficacy of an REP protocol with children with ASD (i.e., Kilroe, Murphy, D Barnes-Holmes & Y Barnes-Holmes, 2014). While similar in many respects, Study 6 advanced on this previous investigation in two ways. First, the aim, in common with
other studies in this thesis, was to examine REP relevant responding itself and in particular the ability to evaluate relational coherence by responding either “Yes” or “No”. As such, participants with a less advanced repertoire than those in Kilroe et al. were included. Second, participants performance on the the NSD-REP were correlated with a standardised language measure (i.e., the PLS4) at pre and post-test in an effort to examine the effects of REP training on intellectual potential.

The results of Study 6 provide some useful preliminary data concerning the assessment and training of foundational relational repertoires in children with ASD using an REP. First, analysis of participants’ PLS4 scores showed a strong positive correlation between language capacity and relational abilities. Second, three children who passed NSD-REP Level 1 but failed Level 2 were successfully trained on the latter using a concurrent MBD. Third, in all three cases, the participants showed a substantial rise in their age equivalent PLS4 score. Again, this pattern of findings, even the speculative ones suggested in the last point, adds to increasing evidence indicating the importance of relational responding for human language and the possible benefits of using the REP to train young children, including those with neurodevelopmental disorders.

Studies 7-9

Based on the outcomes of the studies presented in Chapters 6 and 7, additional research questions were identified. These included: (i) whether the (relatively basic) NSD-REP protocol might be a suitable measure to employ with slightly older children (i.e., 6-7 years) for purposes of producing similar patterns of data; and (2) whether and to what extent NSD-REP Level 3, which featured negation controlled responding, might be trained.
The empirical work presented in Chapter 8 furthered earlier REP efforts by correlating the NSD-REP protocol, not just with standardised assessments of intellectual ability (i.e., the SB5) as had been done in the previous studies, but with the Relational Abilities Test (RAI; Cassidy et al., 2011), a measure of relational framing ability, and tests of educational attainment (i.e., the Drumcondra Primary Reading Test-Revised [DPRT-R] and Drumcondra Primary Math Test-Revised [DPMT-R]). Additionally, Studies 8 and 9 involved training negation controlled responding (i.e., NSD-REP Level 3) in children aged 4-7 years, which as yet, has been the subject of relatively little RFT empirical work.

Study 7 was similar to Study 1 (Chapter 6) in that performance on the NSD-REP was compared with performance on measures of intellectual potential in groups of comparable size. However, as indicated, Study 7 featured several additional standardised assessments (i.e., RAI, DPMT-R, DPRT-R) which could be administered based on the age range of the children. The findings from this investigation showed a significant correlation between performance on the NSD-REP, the SB5 and the Drumcondra Math test, further establishing its utility as an indicator of intellectual potential. Meantime, the RAI assessment, a derivation of the REP, did not correlate with any of the other assessments. A possible explanation for this finding is that the NSD-REP assesses a range of relational responding at a considerably less advanced level than the RAI, and this likely makes it a more suitable assessment instrument with younger age groups (those tested in the present study were from 6-7) than those previously tested with the RAI (those tested in published RAI work tend to be from 8.5 years up).

In Study 8, three seven year olds were successfully trained to respond to Level 3, without adjustment of the protocol, indicating that negation controlled
General Discussion

responding was probably already a skill within their repertoire. Following this success, Study 9 attempted to train this pattern of responding in three 4 year olds. An adjusted NSD-REP protocol (i.e., Level 3 was divided into three component parts) was used to successfully train up the focal response pattern in these children. Together, the results from Studies 8 and 9 demonstrated that the NSD-REP could be successfully employed to teach negation controlled REP responding (in this case, affirmation and disconfirmation with respect to non-arbitrary same and different relations).

Moreover, the findings of Studies 1 through 9 illustrate the utility of the REP generally, and of the NSD-REP specifically, as both an assessment tool and training platform in children from 2.5-7 years old. This work has implications with respect to both research and practice, which will be discussed later in this chapter. Before addressing such matters in more detail, a recap of the final study in the thesis will be provided.

Study 10

The last study in this research programme also investigated the use of an REP protocol with children. However, this study differed in a number of ways from those presented in Chapters 6-8. First, it employed a different, more sophisticated version of the REP (i.e., the SMART), which focused on a greater range of relational ability, including both arbitrarily applicable as well as non-arbitrary relations. Second, this study was employed to further advance a previously instigated research agenda (see Cassidy et al., 2011). Third, the age range of the children that participated in this research, similar to previous SMART-protocol work, was in the pre-teen range (i.e., 8.5-12 years).
In order to address some of the limitations cited in the Cassidy et al. (2011) study, the final empirical investigation explored the effect of derived relations training under more controlled conditions. First, it featured a substantially larger sample of children (n = 28). Second, it included an active control condition who received exposure to an intellectually demanding training regime (in “Scratch” computer coding) for the same duration as the experimental condition received DRR training (i.e., SMART training). Finally, participants were assessed before and after training, not just on IQ tests as in Cassidy et al., but also on academically relevant dimensions of mathematical and linguistic (reading, writing, and oral) ability.

One of the results from Study 10 was that all children trained utilising SMART, showed significant improvements across various measures of intellectual ability and academic attainment compared with controls. Examination of change from pre- to post-intervention without the consideration of condition showed the improvement for relational ability, for full-scale IQ tests, for each of the sub-tests, and for each of the mathematical and linguistic sub-tests in the other measures. Further, examination of time-by-condition interaction effects showed improvements for SMART training over controls for relational ability, for full-scale IQ, for most of the IQ sub-tests, and for mathematical performance.

Aside from the above pattern, and further supporting the key RFT idea of the importance of relational responding (and in particular DRR) for intellectual potential, solid correlations were also found between relational ability and performance on each of the other assessments at pre-test across all participants. Overall, the outcomes from the concluding investigation constitute strong evidence
that REP-based training of relational framing, and SMART-based training in particular, can enhance children’s intellectual potential.

Having thus reviewed the empirical findings of the thesis, next a number of theoretical implications and issues as well as directions for future research will be explored.

**Theoretical Issues, Implications, and Directions for Future Research**

Across ten empirical studies, the utility of the REP as a methodology for assessing and training relational responding abilities in children was investigated. This research offers a number of strengths over previous work in this domain. At the same time, there are also various limitations with the work presented, which in turn highlight a number of potential avenues for further investigation. This section explores the theoretical implications of this work, in addition to directions for impending empirical investigations under the following sub-headings: (i) extension of RFT work; (ii) training REP performance; (iii) negation; (iv) relational training and intellectual performance; (v) the aptness of both REP protocols for differing age groups; (vi) using a computer-based format; (vi) the utility of the NSD-REP; and (viii) methodological issues.

**Extension of RFT work**

First, the work presented in this thesis provides additional support for the RFT position that relational responding is integral to linguistic and cognitive ability, by demonstrating this relationship across five studies (i.e., 1, 5, 6, 7, 10). Though there is a vast quantity of data already supporting this standpoint (see for example, Dymond & Roche, 2013; Hughes & Barnes-Holmes, 2016), the current data provides supplemental evidence of this link. It does so by correlating two REP formats (i.e., the NSD-REP and SMART) with standardised measures of intellectual
and academic attainment in neurotypical children and children with a diagnosis of ASD, ranging from 2 to 12 years.

With specific reference to the NSD-REP, three studies in the current programme of research (i.e., Studies 1, 6 and 7) compared capacity to perform on this REP format with standardised measures of language and cognition. For instance, in Study 1, a correlational study involving a sample size of 26, NSD-REP performance was shown to correlate highly with performance on the SB5 ($\rho = .82$) and the PLS4 ($\rho = .87$). In Study 6, nine children with ASD were assessed on this protocol, as well as on a standardized test of language (i.e., the PLS4). The results indicated a strong positive correlation ($\rho = .82$) between the two. Finally in Study 7, 23 children were assessed on the NSD-REP as well as on several measures of intellectual potential including the SB5 (an intelligence test), the RAI (a measure of relational framing), the DPRT-R (a standardised reading measure) and DPMT-R (a standardised maths measure). The results once again illustrated positive correlations between the NSD-REP and intellectual measures, this time including the SB5 ($\rho = .67$) and the DPMT-R ($\rho = .54$).

Relating to the SMART (i.e., Study 10), RAI performance at pre-test was shown to correlate with performance on various standardized measures of language and cognition in a sample size of 28. That is, it correlated with one sub-test (i.e., block design) from the WASI (Wechsler Abbreviated Scale of Intelligence), three sub-tests (i.e., reading, numerical operations and spelling) from the WIAT-II (Wechsler Individual Achievement Test Second UK Edition) and one sub-test (i.e., letter/number sequencing) from the WISC-IV (Wechsler Intelligence Scale for Children 4\textsuperscript{th} Edition). Further, the RAI was also associated with standardized school reading (i.e., DPRT-R) and math (DPMT-R) assessments. Collectively, the
results from the aforementioned studies provide additional support of the core relationship between DRR and intellectual performance.

In order to further explore these associations, a suggested direction for future work would be to include a wider range of derived relational skills. This applies most obviously to the NSD-REP work with younger children. In this case, assessment and training were focused solely the evaluation of non-arbitrary same and difference relations. There was, of course, a rationale for this choice, as the work in this thesis was focused on examining REP performance per se and thus a repertoire already in place was preferred as the foundation for the work. Even with this relatively narrow measure, strong correlations were seen. Based on this success however, in future, more theoretically expansive work, incorporating a wider scope of REP-based relational responding, might be explored in similar populations to those employed in this thesis. This may include a broader sweep of non-arbitrary relations as well as some foundational arbitrary relations, the use of which would likely help enhance correlations with intellectual measures.

The SMART online program already examines a larger range of relations than the NSD-REP (i.e., “Same/ Opposite” and “More/ Less”), including non-arbitrary and arbitrary responding. Even in this case, however, though obviously to a lesser extent, a wider range of relations might be incorporated so as to improve the protocol, including, for example, temporal (before–after), which has been explicitly linked with intellectual performance (e.g., O’Hora, Pelaez, Barnes-Holmes, Rae, Robinson, & Chaudhary, 2008), as well as hierarchical, deictic (perspective-taking), and analogical relations, which are more complex patterns than those utilised in the current protocol.
Aside from correlating REP scores with measures of linguistic and cognitive ability, this thesis also extends the current work in RFT literature as it was the first to assess and train relational responding using an REP protocol across a broad spectrum of ages in children, ranging from 2.5 to 12 years. As aforementioned, REP research has mostly involved the training of DRR in adults (e.g., Stewart et al., 2004; Dymond & Whelan, 2010; Ronspies, Schmidt, Melnikova, Krumova, Zolfagari, & Banseet, 2015; Roddy, Stewart, & Barnes-Holmes, 2010). REP research with children, both neurotypical (e.g., Rabelo, Bortoloti, & Souza, 2014; Cassidy et al., 2011) and those with developmental delay (e.g., Walsh, Horgan, May, Dymond & Whelan, 2014) is relatively novel and has predominantly centred on establishing this repertoire in older children. This thesis was the first to examine the potential of the REP with very young children, as young as 2.5 years (see Chapter 6). It also examined when ability to perform on an REP like interface might begin to emerge (see Chapter 6), whether, if children have deficits in REP performance at particular ages (see Chapters 6-9), or given particular forms of developmental delay (see Chapter 8), they might be trained in using it, and assuming that a particular population can perform on an REP interface, how closely their performance might correlate with measured intellectual proclivity, as detailed above.

As well as extending the RFT empirical base, an alternative feature of this research was training the capacity to perform on an REP.

Training REP performance

Rather than training particular patterns of relational responding (relational framing) such as “Same”, “Different”, “Opposite” etc., a core element of this work was training REP performance itself. All of the previous studies utilising the REP methodology with children have focused on training multiple stimulus relations. For
instance, Kilroe, et al., (2014) adapted the IRAP computer software program into an interactive teaching tool (T-IRAP) (‘T’ for ‘Teaching’) to successfully target various relational frames (e.g., comparison) in with ASD. Unlike the aforementioned study, however, the studies presented in this thesis are the first known set of experiential investigations to target training of REP performance itself.

As has been laid out, non-arbitrary “Same” and “Different” relations were chosen as the focus of relational evaluation in most of the studies because the repertoire is relatively basic and it was anticipated, that most of the children would already have this skill, or at least be trained in it. If it can be assumed that a child already has this repertoire, then failure at any level illustrates that the particular expertise required for the NSD-REP are absent and not the ability to do the relational responding being evaluated in the NSD-REP. Essentially the skills assessed and trained in the NSD-REP are arbitrary contextual control over non-arbitrary relational responding (Level 1) second order arbitrary contextual control over non-arbitrary relational responding (Level 2) and negation of second order arbitrary contextual control over non-arbitrary relational responding (Level 3). These capabilities are required for a child to perform correctly on this version of the REP and it is useful to have information on children’s level of performance because this is also important for the REP more generally.

The protocol used in Experiments 1-9 might have been improved in at least one way, by analysing the skills required for Level 1 in particular studies. For example, there are two sub-skills that are needed to pass Level 1. These are non-arbitrary same-different responding and the ability come to under the control of the textual stimuli “Same” and “Different” as cues for sameness and difference responding. If children did not pass Level 1, then assessing and if need be, training
these sub-skills separately is an obvious next step. The second of these subskills can itself be broken down further into ability to come under non-arbitrary contextual control over same – different relational responding and discrimination of the stimuli “Same” and “Different” as potential vehicles for such contextual control. These skills might also be examined separately in further work. The reason the current work concentrated on the three levels of the NSD-REP alone was to allow sufficient time and opportunity for assessment and training research across a range of age groups. However, work focusing in on the sub-skills underlying Level 1 in particular would certainly be important as a future research direction.

Another important direction for future work will be to examine how young children perform on an REP focused on arbitrarily applicable relational responding. This has been the focus of REP research programmes including the SMART (e.g., Tirus, Starbrink, & Jansson, 2016), the RCP (Walsh et al., 2014) and the IRAP or T-IRAP (Kilroe et al., 2014). Assessing and training this skill is the key advantage of using the REP with young children. Though the core point of the work in this thesis was to advance the investigation of the REP per se in children of different ages and in particular in very young and neuro-developmentally delayed children, it will be important to extend this research by examining DRR in similar populations using the REP. Also critical, will be the examination of patterns of relational framing in themselves, as well as their developmental order and how they interact (for example, whether training up one particular frame affects performance in others according to a predictable pattern). As outlined in Chapter 2, relational framing is particularly important as the core process underlying language and cognition. Hence, effective assessment and training of this repertoire is critical for purposes of research and practice and REP-based work in young children has much promise in this regard.
In this respect, it should also be noted, training and testing of non-arbitrary relations is still important. This is because non-arbitrary relations are a key precursor to arbitrarily applicable relations (e.g., Y. Barnes-Holmes, D. Barnes-Holmes, Smeets, Strand, & Friman, 2004; Berens & Hayes, 2007). To this end, upcoming work might also examine the use of the REP for assessing and training non-arbitrary relations in very young children, both as a means of supporting the assessment and training of arbitrarily applicable relations, as well as for investigation of the repertoire of non-arbitrary relational responding itself. The latter, of course, continues to be a critical repertoire even after the emergence of relational framing as it can be relatively intricated, depending on the contextual control involved. For example, forms of contextual control over non-arbitrary relations other than those used in the current protocol might vary with respect not just to relation type (e.g., same, different, more, less) but also physical dimensions (shape, colour etc.). As such, quite complex forms of non-arbitrary relational responding might be assessed and trained using the REP and such training could prove very important in supporting the emergence of relational framing.

Aside from training REP performance, another key aspect of this work was the examination of negation controlled responding in children.

**Negation**

Negation is an important aspect of a relational repertoire that children need to acquire. In Chapter 8, the results from the mainstream psychological literature suggested that there appeared to be delayed comprehension of negative sentences compared to positive (or unnegated) sentences (e.g., Clark & Chase, 1972; Just & Carpenter, 1971). The results from Studies 8 and 9 might be seen as showing similar results in a broad sense, given that the relatively older children (i.e., 6-7 years)
involved in these investigations performed well on Level 3 ("Not Same/Not Different"), requiring only three to six sessions to reach criterion in the intervention phase. Conversely, the younger children (i.e., 4 years), despite doing reasonably well on lower levels, needed a remedial type of multi-phase training and required between eleven and fifteen training sessions to complete training. Regarding the remedial multi-phase training, Level 3 had to be separated into three component parts ("Same/Not Same"; "Different/Not Different"; "Not Same/Not Different") to support the younger participants in their attainment of the relevant criteria. As previously alluded to, perhaps the younger children had received insufficient MET opportunities to come under appropriate negation-type control. It is also possible that their performance was a function of the additional complexity of the contextual control involved in tests such as Level 3 of the NSD-REP. That is, they may be less likely to come under appropriate control either due to less experience of language/relational responding in general or of complexity per se.

Although negation controlled responding was trained in Studies 8 and 9, these investigations did not focus on negation per se. Hence, this is something that future REP work could do. For example, one obvious direction might be to investigate and possibly train contextual control that flips between negation and lack of negation. The closest to an empirical demonstration of this kind of pattern thus far seen in RFT research with young children was provided by O’Connor, Barnes-Holmes, and Barnes-Holmes (2011) who investigated relational responding in typically developing children and children with autism. In this study, contextual control was established for emitting either what was referred to as the symmetrical response (i.e., trained A1-B1, tested for B1-A1), or for emitting the alternative response, which was termed the asymmetrical response (i.e., trained A1-B1, tested
for B2-A1). While the cues used in this protocol were, accordingly, labelled in terms of symmetry and asymmetry, they might also be interpreted as “Same” and “Not Same” respectively (for a detailed description of this study, please see p. 36-37).

The switching of contextual control is important for flexibility in responding (e.g., Healy, Barnes-Holmes & Smeets, 2000). As previously alluded to, one of the reasons described for researching the REP with children is that this format has given rise to various protocols that have been used to assess and/or train potentially important examples/properties of relational responding in adults which we might be interested to also examine in children. One of those mentioned in the introduction, for example, was the IRAP, which is an example of the REP in which switching of contextual control is deliberately employed for various purposes, including tapping into previous history of relational responding (e.g., Vahey Barnes-Holmes, Barnes-Holmes & Stewart, 2009). Indeed, in Chapter 8, it was suggested that it was important to examine responding under control of negation cues because such cues were common in applications of the IRAP. Thus, in order to fully test capacity for IRAP performance, such switching should also be investigated in children as it may positively influence intellectual potential.

For instance, O’Toole & Barnes-Holmes (2009) found that performance on an IRAP designed to assess participants’ fluency in DRR correlated with intelligence quotients (IQ) as measured by the Kaufman Brief Intelligence Test. Specifically, undergraduate students were presented with “Same/Different” and “Before/After” tasks in which they were to answer “True” or “False” to pairs of stimuli congruent with a particular relation, such as answering “True” when presented with “table-different-cat.” They were also presented with tasks in which correct answers were incongruent with the stated relation—such as answering “True” when presented with
“table-different-desk.” Results showed that latencies for both types of “same/different” tasks as well as difference (flexibility) scores were correlated strongly with IQ scores, with difference (flexibility) scores for both types of relations strongly correlated with the verbal subtest in particular, suggesting that training relational flexibility, might be critical for improving language skills and promoting intelligent behaviour.

Hence, one eminently possible advance from the current research with the NSD-REP protocol might involve adding a fourth level of difficulty to the current three levels by combining Levels 2 (i.e., “Yes” or “No” in response to a pair of identical or non-identical pictures presented together with an auditory recording of the word “Same” or “Different”) and 3 (which is similar to the latter level but increases task complexity by requiring control by the contextual cue “Not”) and presenting tasks in quasi-random order so that negated and non-negated questions of various types may be equally presented. Participants would then be required to flip between negation-controlled and non-negation-controlled responding.

Future research with the SMART online program, may also benefit from incorporating additional “negation type” contextual control into the paradigm. Such work might involve simply including “Not” or “False” in particular questions or by requiring participants to choose all the available wrong answers as opposed to correct one (O’Connor et al., 2011).

**Relational Training and Intellectual Performance**

In addition to employing progressive protocols to assess and train relational responding, an alternative feature of this work is that it also supports previous RFT research that relational training can boost intellectual performance. With specific reference to the NSD-REP, Studies 5 and 6 documented increased scores on
standardised language and cognition assessments post training. For example, in Study 5, participants tended to show improvement relative to their ages on the PLS4 and the SB5. For the PLS4, all participants except 1 showed improvement while for the SB5, eight out of twelve did so. In Study 6, in all three cases, the participants showed a substantial rise in their age equivalent PLS4 score, post-intervention. This pattern of findings arguably adds to evidence of the importance of relational responding for human language and cognition. Yet, the evidence from these studies was tentative at best and thus further work is certainly needed. One reason for the scepticism is that the kind of training provided in all cases in the current research programme, with the exception of the final study, would not be expected to be sufficient to produce substantial changes in intellectual potential. This is something that needs better controlled investigations with versions of the REP more sophisticated and complex than the NSD-REP, while simpler than SMART; in other words, designed, based on some of the empirical data and theoretical discussion provided here, for the types of populations involved in the current programme.

The data presented in Chapter 9 (i.e., employing the SMART online protocol) represents a much more robust example of the use of relational training to boost intellectual performance in children, albeit in older children with a more advanced relational repertoire. In this respect, this study has also expanded on previous work (e.g., Cassidy et al., 2011), which illustrated that training in DRR positively affected the intellectual potential of children aged 8.5-12 years. The current work showed several advances over Cassidy et al. (2011), including a much larger sample (n=28) of participants, an active control condition and testing before and after the intervention stage, not only on several different IQ type sub-tests but also on academically relevant dimensions of mathematical and linguistic ability.
Similar to the original study, significant post-intervention differences in performance in respect of both mathematical and linguistic skill were detected between the REP and control conditions, further indicating the importance of DRR in increasing intellectual performance.

In order to fully explore this relationship, an important direction for future work will be to use bigger, better controlled studies, conducted over longer periods of time. This is particularly relevant to work with the NSD-REP, though it can arguably be applied to the SMART work also. For instance, the training periods for all studies in the current programme of research were of relatively brief duration. Relating to the NSD-REP, sessions were approximately two to five minutes in length and the number of training sessions ranged from 3 to 53. Similarly, the training in Study 10 (i.e., the SMART study) was conducted once a week over a five month period and totalled just 29 hours. By this point, better controlled studies, conducted over prolonged time periods are required and while they have started to emerge for the SMART protocol (see Cassidy et al., 2016 as an example), these empirical investigations also need to be focused on REP protocols that can be used with younger children.

Also related to this point is the issue of sustainability of the direct effects of relational training itself as well as the possible knock-on beneficial effects on intellectual potential. In studies using the NSD-REP, the ability to respond to non-arbitrary same and different relations under arbitrary contextual control was maintained in the follow-up data. This data was taken two to four weeks post-intervention for several studies in the thesis, including studies 2, 3, 8 and 9. However, not all of the investigations utilising this methodology included follow-up data. Further, none of the aforementioned studies collected follow-up data on the
standardised language and cognitive assessments (e.g., the SB5 or the PLS4).

Regarding the SMART study, the data presented here arguably offer no insight into the sustainability of the effects of DRR training, which is an important issue, given the theoretical argument that DRR is key to intellectual performance. The DRR training in the SMART study clearly enhanced the test performances of the experimental group, but without follow-up tests it cannot be argued that real and lasting change in intellectual ability might have been produced.

While research on the effects of DRR specifically as an intervention for improving intellectual potential is only beginning (see Roche, Cassidy, & Stewart, 2013), there is preliminary evidence to supplement the theoretical position that DRR training, such as used in the final study, may indeed produce long-term improvement in intellectual ability. Nevertheless, additional evidence would further enhance the case, and thus ideally, studies such as the current one should aspire to collect follow-up data on participants’ IQ and intellectual performance more broadly.

**The suitability of the REP protocols for differing age cohorts**

Another issue for comment in the context of the current research programme relates to the utility of both REP protocols. The results of the thesis studies suggest that the NSD-REP and the SMART are both useful relational assessment/training protocols, suitable for different age ranges.

The NSD-REP showed utility with the younger participants involved in this research programme. Across nine investigations, this REP protocol was utilised successfully with neurotypical children and children with developmental delay. Moreover, the results of these studies highlight the potential of the NSD-REP as an educational tool for children ranging from young to middle-childhood.
At the same time, these data also suggest a limit to the age range for which the NSD-REP, in its current form, might be used. Participants in Study 4, who were under the age of three ($M = 2.09$), did not show learning of the Level 1 repertoire without the aid of non-automated or table-top procedures. These participants were exposed to 46 and 72 training sessions respectively prior to attaining criterion, with the second participant (who was three months younger than the first), requiring almost double the number of exposures. Additionally, coherent with previous work on the correlation between relational responding and language/intellectual level, language and cognition age-equivalent scores seemed to correlate with length of time required to demonstrate Level 1 competence. For instance, both participants had an average PLS4 age-equivalent score of 2.04 years and an average SB5-ABIQ age-equivalent score of 2.08 years. In comparison, the children that participated in Study 2, who were also exposed to Level 1 of the NSD-REP, attained criterion in three to five sessions. These participants were over the age of three ($M = 3.04$ years) with average PLS4 age equivalent scores of 3.00 years and SB5-ABIQ age-equivalent score of 3.07 years. This suggests a verbal and cognitive range that predicts success on the NSD-REP. Of course this sample ($n = 6$) is small and hence, the results remain speculative at this time. Further NSD-REP research with children of similar ages and profiles is warranted before firm conclusions might be drawn.

Relating to the SMART online program, previous research (i.e., Cassidy et al., 2011) involved participants aged 8.5-12 years. Similarly, the work presented in Chapter 9 of this thesis involved the testing and training of Same/Opposite and More/Less relations with participants aged 10-11 years. However, an assessment using the RAI (the assessment test corresponding to the SMART, described in Chapter 8) with participants aged 6-7 years indicated that this protocol may not be
suitable for children of this age range. On average, the children scored a total of 27 correct responses out of a possible 55 and scores ranged from 20-34. These scores were relatively lower and showed much less variability in comparison to those seen in Cassidy et al. (i.e., 46.75; 29-60.7) and in Study 10 (i.e., 31; 16-46). Moreover, verbal reports from the children during and after the assessment demonstrated that they found the RAI particularly difficult. As suggested, even though derived relational responding provides the basis for intellectual performance and develops from an early age across the elementary frames assessed in the RAI, several factors may have impacted performance including: (i) the children in this age group have less experience of relations than children in older age groups; (ii) the test may have required a duration of on-task behaviour that was relatively extended for this age group; and/or (iii) the level and/or the range of derived relational responding required may have been unsuitable or too high for the children assessed. For the REP/SMART also, then, a particular verbal and cognitive range predicts success, and for this protocol, the minimum is considerably higher than for the NSD-REP.

Computer Based Format

In order to examine relational performance, all of the studies described had the additional advantage of using computer-based procedures. Previous work investigating relations between derived relations and linguistic or cognitive ability, in such populations, has typically employed table-top procedures. However, of late, computer-based procedures have been utilised for variety of different types of training (e.g., Walsh, Horgan, May, Dymond & Whelan, 2014) as they offer a number of advantages (e.g., Whalen, Liden, Ingersoll, Dallaire, & Liden, 2006). For instance, participants are able to respond at their own pace as opposed to an instructor dictating the pace of instruction (Kilroe et al., 2014; Walsh et al., 2014).
Additionally, the computer can present materials automatically in either pre-designated or deliberately randomised order. As such, sub-optimal features of instructor-based manipulation of materials, including failure to counter balance position of materials, or significantly slow instructional pace, are removed. Further, employing an automated procedure allows for accurate, reliable and valid data collection (Kilroe et al., 2014; Walsh et al., 2014).

Hence, by investigating the assessment and training of very young children in computer-based REP-type relational responding tasks, this work has added to the existing literature in this area, not simply with regard to the investigation of REP training per se in such populations, but also in relation to the use of computer-based teaching in this domain.

The Utility of the NSD-REP

The efficacy of the NSD-REP was demonstrated by the fact that the assessment was invariably rapid (e.g., average assessment time of approximately six minutes) and unproblematic (all children could be tested and scored) as a means of measuring relational performance in young children ranging from 2-7 years. In addition, despite relatively limited training, there is preliminary evidence to suggest that the effects of NSD-REP training are maintained post intervention (as seen in Studies 2, 3, 8, 9).

Given these features, it seems that the NSD-REP may have certain advantages over the SMART. As illustration, the NSD-REP is much simpler and faster to use in terms of assessment at least, yet it shows high levels of correlation with standardised measures (e.g., the SB5, the PLS4, and the DPMT-R). Related to this point is that the NSD-REP protocol significantly correlated with cognitive and school based standardised assessments for the children involved in Study 7, while
the SMART did not. As highlighted in this chapter, the levels of relational responding and on-task behaviour required by the SMART are higher than those required by the NSD-REP. Hence, in the absence of preparatory training, the latter may simply be more suitable as an assessment tool for this age range. With regard to maintenance of training effects, although empirical investigations using the SMART have demonstrated significant increases in intellectual performance post-training (e.g., Cassidy et al., 2011, 2016), there is, as yet, relatively little evidence to demonstrate whether children maintain the ability to respond successfully on this protocol after time has elapsed.

Of course, the NSD-REP focuses on a relatively limited range of relational responding and a narrower range than the SMART. However, if results such as these can be achieved with an REP focused on contextual control over simple non-arbitrary same and difference relations, then expansion to a wider field of non-arbitrary relations and to arbitrary relations seems feasible and is certainly worthy of future research. Such work might help assessment with respect to aspects of relational repertoires already present, as well as helping to establish new repertoires. Autonomously, this protocol may be a useful gauge in future studies, behavioural or otherwise, not just of relational responding, but of intellectual performance more generally.

**Methodological Issues**

While there were a number of strengths featured in the current work (e.g., the utility of the NSD-REP), some methodological issues were also evident and should be addressed in future empirical investigations. First, it could be argued that the administration of all tests (i.e., the REP protocols and alternative linguistic/cognitive measures) by the same researcher is a methodological weakness of this work.
Though the resources needed to implement this control (utilising multiple researchers) were not available in the current research programme, it should be a priority in future studies. Ideally this should involve blind testing also.

In addition, at least some of the standardised assessments used in this thesis required a minimum period between testing and retesting, which was not always adhered to. For example, the PLS4 and SB5 both require six months between exposures (e.g., Bain & Allin, 2005). Yet, the re-administration of these assessments in Studies 5 and 6 was often performed within a period between three and six months after initial testing. Hence, in some circumstances, the minimum requirement/criteria was not met for either instrument. This means a retesting effect may have been seen in these studies, which may partly account for rises seen in participants’ scores. Future research should ensure that an appropriate amount of time has elapsed between testing occasions.

Regarding the NSD-REP specifically, response accuracy data was used to measure performance. Despite the widespread use of this as a dependent variable, Kilroe et al. (2014) cautions that accuracy as a dependent measure may overlook important subtleties in responding. That is, it cannot show any difference between accurate but struggling performance, and fluent performance. Without measuring time, progress toward goals cannot be precisely monitored and/or altered based on individual performance. Further, speed and accuracy of responding are important dimensions tested by the IRAP (e.g., Rabelo, Bortoloti & Souza, 2014) and the SMART (Cassidy et al., 2011). Hence, another future direction presented for this protocol is the use of speed and accuracy data presented by the computer (Kilroe et al., 2014; Wilkinson, 2008).
General Discussion

Relating to the SMART protocol, an alternative methodological limitation pertains to the difference in performance across participants. While some participants could complete all SMART levels (i.e., 70 levels) involved at least once, others struggled to complete even just a few. The latter repeatedly failed particular stages (e.g., 21 and 28 which involved responding to complex questions relating to “Same” and “Opposite” relations) and as a result showed waning enthusiasm for the training. Although a number of environmental manipulations (e.g., customized reinforcement strategies) were used, they were not enough to maintain motivation/ performance in some children. Future work should aim to improve the reinforcing value of the tasks and address possible deficits in relational ability that might otherwise hamper progress and enthusiasm.

Summary and Conclusion

To summarise, the overarching goal of the current program of research was to investigate the REP as a protocol as an assessment and training instrument in children across a spectrum of ages. Employing two alternative procedures, namely the NSD-REP and the SMART, the results presented in this thesis have been encouraging.

Relating specifically to the NSD-REP, across four studies this protocol was found to correlate consistently with both linguistic and cognitive ability. It was also used successfully as a training tool for children with and without disabilities (i.e., Studies 2-4, 6, 8, 9). At a basic level, using the NSD-REP has also provided an insight into the age at which children can begin to respond to an REP interface and how adapting REP protocols (e.g., breaking levels into their component parts) can facilitate attainment on various level requirements.
With reference to the SMART online protocol, the research (i.e., Study 10) provided in this thesis extended a previously initiated research agenda (e.g., Cassidy et al., 2011), by exploring the effects of this online programme under more controlled conditions (i.e., limited training hours, an active control condition, additional standardised tests). Significant improvements were seen in measures of overall intellectual performance and in academic attainment, in the SMART group alone, and there was significantly more improvement on these measures in the SMART group than in the control group. Collectively, the results of Studies 1-10 illustrate the utility of the REP as an assessment, training and potential educational tool for children of various ages.

Aside from providing a recap on the research undertaken, a number of strengths, theoretical implications and possible directions for future studies were also outlined. For instance, some of the strengths highlighted in the previous section included additional support for the link between relational training and intellectual performance, the utility of the NSD-REP as an REP format and the exploration of negation controlled responding. Avenues for further investigation included: (i) expanding the number of relations to be trained and tested, including both non-arbitrary and arbitrary relational responding; (ii) examining the effects of longer term more substantive training utilizing more subtle dependent measures of relational responding (e.g., fluency); and (iii) collecting follow-up data on participants’ IQ and intellectual performance more broadly.

In conclusion, the work presented in this thesis is promising for the potential of the REP as an instrument for use in children and gives an early indication that the protocol could eventually provide the basis of a comprehensive tool for assessing
and training of relational responding with typically developing children and children with developmental delay across a wide spectrum of ages and abilities.


References


References


References


References


doi:10.1007/BF01531712


References


References


References


References


Symposium presented at the 36th Annual Convention of the Association for Behavior Analysis, San Antonio, Texas.


References

Completion Procedure. *The Psychological Record*, 63, 751-768. doi: 10.11133/j.tpr.2013.63.4.004


References


Rabelo, L. Z., Bortoloti, R., & Souza, D. H. (2014). Dolls are for girls and not for boys: evaluating the appropriateness of the Implicit Relational Assessment Procedure for
References


http://www.academia.edu/287331/Grammar_and_Social_Organization_Yes_No_Interrogatives_and_the_Structure_of.Responding


self in typically-developing children and children with dyslexia or ADHD.  

*Behavioral Development Bulletin* 19 (2), 48-59. Retrieved from: 


doi: 10.1186/2050-7283-1-12


29, 137-155. Retrieved from:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3659503/


https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3139559/


References


Appendix A: Samples of the NSD-REP

Opening screen of the NSD-REP. The researcher selects the appropriate level and presses the continue button.
Samples of Level 1 of the NSD-REP

Are these the same or different?

SAME    DIFFERENT

Are these the same or different?

SAME    DIFFERENT
Samples of Level 2 of the NSD-REP

Are these the same?

Are these different?
Samples of Level 3 of the NSD-REP

Are these not the same?

Are these not different?
## Appendix B. Yes/No Prescreening Assessment

Vocal Response Targets and Antecedent Stimuli Presented Under Tact Conditions for a Representative Participant

<table>
<thead>
<tr>
<th>Vocal response targets</th>
<th>Antecedent verbal stimulus</th>
<th>Antecedent nonverbal stimulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tact ‘‘yes’’</td>
<td>‘‘Is this a cup?’’</td>
<td>Picture of a cup</td>
</tr>
<tr>
<td></td>
<td>‘‘Is this a boat?’’</td>
<td>Picture of a boat</td>
</tr>
<tr>
<td></td>
<td>‘‘Is this a bear?’’</td>
<td>Picture of a bear</td>
</tr>
<tr>
<td></td>
<td>‘‘Is this a car?’’</td>
<td>Picture of a car</td>
</tr>
<tr>
<td></td>
<td>‘‘Is this a cow?’’</td>
<td>Picture of a cow</td>
</tr>
<tr>
<td>Tact ‘‘no’’</td>
<td>‘‘Is this a shoe?’’</td>
<td>Picture of a boat</td>
</tr>
<tr>
<td></td>
<td>‘‘Is this a pen?’’</td>
<td>Picture of a bear</td>
</tr>
<tr>
<td></td>
<td>‘‘Is this a ball?’’</td>
<td>Picture of a cup</td>
</tr>
<tr>
<td></td>
<td>‘‘Is this a duck?’’</td>
<td>Picture of a car</td>
</tr>
<tr>
<td></td>
<td>‘‘Is this a house?’’</td>
<td>Picture of a cow</td>
</tr>
</tbody>
</table>
Appendix C: RAI/SMART Test Instructions

Welcome to SMART brain training. The SMART IQ test has been developed and validated by psychologists to assess your IQ and fluid intelligence before you begin your training. During training and at the end of your course, you can retake the test to measure IQ gains.

Instructions:
In a moment, some statements will appear on this screen. The statements will involve made-up words (e.g., ‘PEF’). A question will also be presented underneath these statements. This question will ask you about the relationship between the made-up words. You can work out the answer by reading the statements really carefully. You can answer the questions by clicking on the words ‘yes’ or ‘no’ on the computer screen. It is important that you work out each answer in your head. You have only 30 s in which to answer each question.

DO NOT WRITE ANYTHING DOWN AND DO NOT HAVE ANYONE HELP YOU.

PLEASE ENSURE THAT YOU HAVE NO INTERRUPTIONS FOR THE NEXT 15 MINUTES OR SO.

IF YOU CHEAT ON THIS TEST, YOUR ESTIMATED IQ RISE THAT WILL BE PROVIDED AT THE END OF THE SMART PROGRAM WILL BE UNRELIABLE.

Start Assessment