THE INSTANTIATION OF THEORY INTO PRACTICE FOR PRE-SERVICE TEACHERS IN INCLUSIVE EDUCATION

JULIE ANNE LANCASTER

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at Charles Sturt University

2016
Certificate of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma at Charles Sturt University or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by colleagues with whom I have worked at Charles Sturt University or anywhere else during my candidature is fully acknowledged.

I agree that this thesis be accessible for the purpose of study and research in accordance with the normal conditions established by the Executive Director, Library Services or nominee, for the care, loan and reproduction of theses.

Signature:                      Date:
I would like to express my sincere thanks to the many people who provided encouragement and support during the planning, execution and final writing of this research, and in particular, the following individuals:

Associate Professor Alan Bain as my Principal supervisor for your uncanny ability to always be encouraging, supportive and able to provide prompt feedback.

Dr Sarah McDonagh for her attention to detail and persistence until I ‘got’ the stats.

My daughter Emma who is the most important person in my life. Emma, you taught me the importance of everything.

My parents, John and Gwenda, for always being encouraging and supportive throughout my whole life. You taught me the value of education and of always striving to do my best.

My partner Iain, for being there when I needed you.

And to the many friends and colleagues, across my adult years, in all settings for their support, encouragement and collaboration.

Special thanks must also go to Kim Woodland for her Professional Editorial Assistance which included formatting, grammar and style. No substantive content or conceptual organisation of the thesis was carried out.
Abstract

Pre-service teacher (PST) education courses are the vehicles for providing teachers with the preparation they require to work in inclusive classrooms. Discussions in the literature surrounding teacher preparation for inclusion are extensive (Jung, 2007; Killoran et al., 2014; Romi & Leyser, 2006). While there is widespread support for university-based teacher preparation, there continues to be national and international concern about whether the preparation PSTs receive for inclusion is adequate, and whether it does what schools require to ensure successful inclusive education (AITSL, 2015; Carroll et al., 2003; Edelen-Smith et al., 1993; Forlin et al., 2015; Husebo, 2012; Snyder, 2012). Serious concerns exist about the preparation of teachers for inclusive classrooms and also a lack of knowledge about how this preparation can be carried out differently and better (Cochran-Smith, Villegas, Abrams, Chavez Mills, & Stern, 2015; Darling-Hammond, 2005; Dempsey, 2012b; Florian, 2011).

The research described here investigates whether the application of a course design approach derived from theories of self-organisation and applied to course development influences pre-service teacher mastery of pedagogical content knowledge in inclusive education, professional pattern language, performance on assessments of self-efficacy, and the classroom practice of pre-service teachers.

Quasi-experimental methods were utilised in two studies employed in this research. A non-equivalent comparison group design was used to establish the differential effects of course design, one based on the theoretical principle of embedded design and the other on an approach based upon applied classroom experience. The dependent variables were pedagogical content knowledge, professional pattern language, and self-efficacy. The second study sought to establish differential impacts of course design on the translation of the skills learned in the teacher preparation into practice whilst pre-service teachers were on practicum.

The results indicated statistically significant findings in favour of the embedded design cohort for pedagogical content knowledge and pattern language skills indicating that embedded design may be a promising design approach for providing students with higher levels of important pedagogical content knowledge and the language required to share professional knowledge. This was not the case for levels of self-efficacy where
differences between the cohorts were not significant although PSTs in both cohorts improved in self-efficacy at statistically significant levels over the course of the study. The results of data collected in study two including classroom observations, PST reflections, and an evaluation of lesson plans indicated positive effects in favour of the *embedded design* cohort for design, implementation, and also self-reflection about the fidelity of lesson design and delivery when the skills were applied in a practicum setting. Recommendations are made for the design of pre-service teacher inclusive education courses.

N.B. throughout the thesis the term ‘course’ is used to refer to a single unit (subject) of study for PSTs within their pre-service teacher education program.
# Table of Contents

Certificate of Authorship ........................................................................................................ i

Acknowledgements .................................................................................................................. ii

Abstract .................................................................................................................................. iii

Table of Contents .................................................................................................................... v

List of Tables .......................................................................................................................... xiii

List of Figures .......................................................................................................................... xv

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

1.1 The Importance of Inclusion and Inclusive Teacher Practice in Schools ........................................ 1

1.2 The Importance of Inclusion and Inclusive Teacher Practice in Pre-service Teacher Education Programs .................................................................................................................. 3

1.3 Evaluating the Courses Used by Faculties to Prepare Pre-service Teachers for Inclusion .......................................................... 5

1.4 Theoretical Frameworks Evident in Pre-service Teacher Education Programs ........................................... 7

1.5 Summary ............................................................................................................................ 10

Chapter 2: Review of the Literature ......................................................................................... 12

2.1 Inclusive Teacher Practice Context ..................................................................................... 13

2.1.1 A brief introduction about the nature of inclusion ............................................................. 13

2.1.2 The history leading to current practices ........................................................................... 14

2.1.3 The importance of quality inclusive teacher practice ....................................................... 16

2.1.4 What do teachers think about inclusion and why does it remain a challenge? ......................... 17

2.1.5 What do teachers need to know to make inclusion happen? ............................................ 17

2.2 Teacher Use of Research-Based Pedagogies Required to Enhance Inclusion .............................................. 18

2.2.1 Explicit teaching ............................................................................................................... 19

2.2.1.1 Explicit teaching description ...................................................................................... 19

2.2.1.2 Meta analyses that provide a rationale for the use of explicit teaching ............................ 19

2.2.1.3 Specific research studies that lend support for explicit teaching ................................. 22

2.2.1.4 Issues and challenges associated with explicit teaching ........................................ 24
Chapter 3: Method

3.1 Study One

3.2 Research Questions

3.3 Hypotheses

3.4 Research Design

3.4.1 Threats to validity and control over experimental designs

3.4.1.1 Internal validity

3.4.1.2 Statistical conclusion validity

3.4.1.3 External validity

3.5 Threats to Internal Validity

3.5.1 Selection bias

3.5.2 History

3.5.3 Maturation

3.5.4 Attrition/Mortality

3.5.5 Additive and interactive effects to validity

3.5.6 Testing

3.5.7 Instrumentation

3.5.8 Statistical regression

3.6 Threats to Statistical Conclusion Validity

3.6.1 Low statistical power

3.6.2 Violated assumptions of statistical tests

3.6.3 Error rate problem

3.6.4 Reliability of measures

3.6.5 Reliability of treatment implementation

3.6.6 Extraneous variance in the experimentation setting

3.6.7 Heterogeneity of respondents

3.7 Threats to Construct Validity

3.7.1 Inadequate operational explication of constructs: Independent variables

3.7.2 Inadequate operation of the dependent variables

3.7.3 The Hawthorne effect

3.7.4 Experimenter effects
3.7.5 Confounding constructs and level of constructs. ........................................ 129

3.8 STUDY ONE PROCEDURES ........................................................................ 130

3.8.1 Participants ......................................................................................... 130
  3.8.1.1 Instructors .................................................................................... 131

3.8.2 Conditions. .......................................................................................... 132

3.8.3 Embedded design and levels of the independent variable. ............... 133
  3.8.3.1 Self-organizing systems theory. ..................................................... 133
  3.8.3.2 Embedded design. ....................................................................... 134

3.8.3.3 Component 1: Knowledge and awareness (Level 1 of embedded
design). ................................................................................................. 134

3.8.3.4 Component 2: Active experience for embedded design .......... 135

3.8.3.5 Component 3: Continuous feedback for embedded design ...... 135

3.8.3.6 Component 4: Personal impact for embedded design ............ 136

3.9 THREATS TO EXTERNAL VALIDITY .................................................... 136

3.9.1 Lack of representativeness of available populations/selection-treatment
interaction. ................................................................................................. 137

3.9.2 Interaction of setting and treatment. .................................................. 138

3.9.3 Interaction of history and treatment. ................................................. 138

3.9.4 Teaching cycle for the embedded design cohort. .......................... 138

3.9.5 Conditions for the applied experience cohort. .............................. 139

  3.9.5.1 Applied experience as an independent variable. Level I: Knowledge
and awareness. ...................................................................................... 140

  3.9.5.2 In-situ experiences. ....................................................................... 141

3.9.6 Teaching cycle for the applied experience cohort. ....................... 142

3.9.7 Reliability of instructor implementation ........................................ 144

3.10 DATA COLLECTION PROCEDURES ................................................. 145

3.10.1 Dependent measures used for study one. ..................................... 146

  3.10.1.1 Self-Efficacy questionnaire (SEIPD). ....................................... 147

  3.10.1.2 Reliability of the SEIPD. ......................................................... 147

  3.10.1.3 Validity of SEIPD. ................................................................. 148

  3.10.2.1 Pedagogical content knowledge (PCK). .................................. 149

  3.10.2.3 Content validity of the pedagogical content knowledge test. .... 150

  3.10.2.2 Reliability of pedagogical content knowledge ......................... 151

  3.10.2.4 Construct validity of the pedagogical content knowledge test..... 152

  3.10.3.1 Pattern language (PL) .......................................................... 153
3.10.3 Frequency of pattern language ..............................................................153
3.10.3.3 Sophistication of pattern language ..................................................155
3.10.3.4 Reliability of pattern language frequency and sophistication ............156
3.10.3.5 Validity of pattern language ...........................................................156
3.11 DATA ANALYSIS ......................................................................................158
3.11.1 Pedagogical content knowledge analysis .............................................158
3.11.2 Pattern language frequency .................................................................159
3.11.3 Pattern language sophistication ............................................................159
3.11.4 SEIPD analysis ......................................................................................162
3.12 Study two ..................................................................................................162
3.12.1 Research questions ...............................................................................162
3.13 HYPOTHESES .........................................................................................163
3.14 RESEARCH DESIGN ................................................................................164
3.14.1 Threats to validity and control over experimental designs ..................165
3.15 THREATS TO INTERNAL VALIDITY .......................................................166
3.15.1 Selection bias .........................................................................................166
3.15.2 History ..................................................................................................167
3.15.3 Maturation .............................................................................................167
3.15.4 Attrition /Mortality ...............................................................................167
3.15.5 Additive and interactive effects to validity ...........................................167
3.15.6 Testing ...................................................................................................167
3.15.7 Instrumentation .....................................................................................168
3.15.8 Statistical regression ............................................................................168
3.16 THREATS TO STATISTICAL CONCLUSION VALIDITY .........................168
3.17 THREATS TO CONSTRUCT VALIDITY ....................................................170
3.18 THREATS TO EXTERNAL VALIDITY ......................................................170
3.19 PROCEDURES ..........................................................................................171
3.19.1 Conditions ............................................................................................171
3.19.3 Participants ............................................................................................171
3.19.3 Procedures for study two .......................................................................173
3.20 DATA COLLECTION PROCEDURES ......................................................174
3.20.1 Dependent measures .............................................................................174
3.20.1.1 Self-Efficacy questionnaire (SEIPD) ..................................................175
3.20.1.2 Lesson Reflection used for pattern language ..................................175
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.20.1.3 Design Fidelity of the Lesson</td>
<td>175</td>
</tr>
<tr>
<td>3.20.1.4 Implementation fidelity</td>
<td>175</td>
</tr>
<tr>
<td>3.20.1.5 Self-reflection fidelity</td>
<td>177</td>
</tr>
<tr>
<td>3.20.2.1 Reliability of the fidelity observations</td>
<td>177</td>
</tr>
<tr>
<td>3.20.2.2 Validity of the fidelity observations</td>
<td>178</td>
</tr>
<tr>
<td>3.21 DATA ANALYSIS</td>
<td>179</td>
</tr>
<tr>
<td>3.21.1 Self-efficacy analysis</td>
<td>179</td>
</tr>
<tr>
<td>3.21.2 Pattern language frequency and sophistication</td>
<td>180</td>
</tr>
<tr>
<td>3.21.3 Design fidelity for explicit teaching</td>
<td>180</td>
</tr>
<tr>
<td>3.21.4 Implementation fidelity for explicit teaching</td>
<td>180</td>
</tr>
<tr>
<td>3.21.5 Self-reflection fidelity for explicit teaching</td>
<td>181</td>
</tr>
<tr>
<td>3.22 SUMMARY</td>
<td>182</td>
</tr>
<tr>
<td>Chapter 4: Results</td>
<td>183</td>
</tr>
<tr>
<td>4.1 STUDY ONE: GROUP EQUIVALENCE AT PRE-TEST</td>
<td>183</td>
</tr>
<tr>
<td>4.1.1 Reliability of instructor implementation</td>
<td>185</td>
</tr>
<tr>
<td>4.2 APPLICATION OF THE EMBEDDED DESIGN PRINCIPLE TO THE DESIGN OF A</td>
<td></td>
</tr>
<tr>
<td>PRE-SERVICE TEACHER EDUCATION INCLUSIVE EDUCATION COURSE</td>
<td>186</td>
</tr>
<tr>
<td>4.3 MASTERY OF COURSE CONTENT IN THE FORM OF PEDAGOGICAL CONTENT</td>
<td>189</td>
</tr>
<tr>
<td>4.4 FREQUENCY AND SOPHISTICATION OF PATTERN LANGUAGE</td>
<td>193</td>
</tr>
<tr>
<td>4.4.1 Frequency of pattern language at pre-post occasion</td>
<td>194</td>
</tr>
<tr>
<td>4.4.2 Frequency of pattern language at quiz occasion</td>
<td>196</td>
</tr>
<tr>
<td>4.4.3 Sophistication of pattern language at quiz occasion</td>
<td>198</td>
</tr>
<tr>
<td>4.5 SELF-EFFICACY</td>
<td>201</td>
</tr>
<tr>
<td>4.6 STUDY TWO RESULTS</td>
<td>202</td>
</tr>
<tr>
<td>4.6.1 Frequency of pattern language</td>
<td>204</td>
</tr>
<tr>
<td>4.6.2 Sophistication of pattern language</td>
<td>205</td>
</tr>
<tr>
<td>4.7 SELF-EFFICACY</td>
<td>206</td>
</tr>
<tr>
<td>4.8 APPLICATION OF THE EMBEDDED DESIGN PRINCIPLE IN PRACTICE TEACHING</td>
<td></td>
</tr>
<tr>
<td>SETTINGS</td>
<td>207</td>
</tr>
<tr>
<td>4.8.1 Dependent variables</td>
<td>207</td>
</tr>
<tr>
<td>4.8.2 Lesson Design Fidelity</td>
<td>210</td>
</tr>
<tr>
<td>4.8.3 Implementation fidelity</td>
<td>211</td>
</tr>
<tr>
<td>4.8.4 Self-evaluation fidelity</td>
<td>214</td>
</tr>
</tbody>
</table>
4.9 SUMMARY ......................................................................................................................... 215

Chapter 5: Discussion ............................................................................................................. 217

5.1 PEDAGOGICAL CONTENT KNOWLEDGE (PCK) ............................................................ 220
  5.1.1 The research question ......................................................................................... 220
  5.1.2 Summary of research findings ........................................................................... 220
  5.1.3 Discussion of the research findings ................................................................. 221

5.2 PATTERN LANGUAGE ..................................................................................................... 222
  5.2.1 The research questions ..................................................................................... 222
  5.2.2 Summary of research findings ........................................................................ 223
  5.2.3 Discussion of the research findings about pattern language ......................... 224

5.3 SELF-EFFICACY (SE) ..................................................................................................... 230
  5.3.1 The research questions ..................................................................................... 231
  5.3.2 Summary of research findings ........................................................................ 231
  5.3.3 Discussion of the results .................................................................................. 232

5.4 DESIGN, IMPLEMENTATION, AND SELF-REFLECTION INTEGRITY OF INCLUSIVE
CLASSROOM PEDAGOGIES .................................................................................................. 236
  5.4.1 The research questions ..................................................................................... 236
  5.4.2 Summary of research findings ........................................................................ 236
  5.4.3 Discussion of the research findings ................................................................. 238
    5.4.3.1 Design fidelity .............................................................................................. 239
    5.4.3.2 Implementation fidelity ............................................................................... 240
    5.4.3.3 Self-evaluation fidelity ............................................................................... 242

5.5 SIMULATED SETTINGS .................................................................................................... 243

5.6. IN-SITU SETTINGS .......................................................................................................... 244

5.7 LINKS BETWEEN A THEORETICALLY DESIGNED COURSE AND PST
PRACTICE .............................................................................................................................. 246

5.8 LIMITATIONS OF THE STUDY .................................................................................... 248
  5.8.1 Design ............................................................................................................... 248
    5.8.1.1 Statistical conclusion validity. ................................................................. 249
    5.8.1.2 History ...................................................................................................... 250
  5.8.2 Reliability of the dependent measures ............................................................ 250

5.9 FUTURE DIRECTIONS ..................................................................................................... 254

5.10 CONCLUSIONS .............................................................................................................. 258

References .............................................................................................................................. 259
Appendices ................................................................................................................................................. 281

Appendix 1: Course implementation data collection tool across cohort delivery .................................................................................................................................................. 282

Appendix 2. Component 1 of embedded design: Knowledge and awareness. ...... 290

Appendix 3: Weekly schedule / teaching cycle of the embedded the design principle throughout the content. ........................................................................................................................................ 292

Appendix 4: Pedagogical content knowledge data collection formats ............. 293

Appendix 5: Decision table for coding of SOLO responses for three quiz occasions .................................................................................................................................................. 294

Appendix 6: Observation measures used for data collection of design and implementation fidelity ........................................................................................................................................ 295

Appendix 7: SEIPD data collection formats ......................................................................................... 296

Appendix 8: Demonstration example of an explicit teaching lesson: ............. 298

Appendix 9: SERAP approval ................................................................................................................. 299
List of Tables

TABLE 2.1 JOURNAL RANKINGS USING DESCRIPTORS “PRE-SERVICE TEACHER EDUCATION AND INCLUSION” ................................................................. 54

TABLE 2.2 REVIEW OF THE RESEARCH-BASED LITERATURE ON PRE-SERVICE TEACHER PREPARATION FOR INCLUSION ................................................................. 57

TABLE 3.1 THE STATIC-GROUP PRE-TEST/POST-TEST DESIGN .................................................................................. 109

TABLE 3.2 ESSENTIAL QUALITY INDICATORS FOR GROUPS AND QUASI-EXPERIMENTAL RESEARCH PROPOSALS ................................................................................. 110

TABLE 3.3 POTENTIAL THREATS TO INTERNAL VALIDITY IN THE CURRENT RESEARCH .................................................. 116

TABLE 3.4 POTENTIAL THREATS TO STATISTICAL CONCLUSION VALIDITY ........................................................................ 122

TABLE 3.5 POTENTIAL THREATS TO CONSTRUCT VALIDITY .................................................................................. 127

TABLE 3.6 THREATS TO EXTERNAL VALIDITY .................................................................................. 137

TABLE 3.7 DEMOGRAPHIC INFORMATION OF PARTICIPANTS IN STUDY ONE ........................................................................ 134

TABLE 3.8 SUMMARY OF DIFFERENCES BETWEEN THE COHORTS .................................................................................. 143

TABLE 3.9 PATTERN LANGUAGE TERMS .................................................................................. 154

TABLE 3.10 SOLO CODING SYSTEM CATEGORIES (BIGGS & COLLIS, 1982) ........................................................................ 155

TABLE 3.11 THE STATIC-GROUP PRE-TEST/POST-TEST DESIGN .................................................................................. 166

TABLE 3.12 POTENTIAL THREATS TO INTERNAL VALIDITY IN THE CURRENT RESEARCH ...................... 163

TABLE 3.13 POTENTIAL THREATS TO STATISTICAL CONCLUSION VALIDITY ........................................................................ 165

TABLE 3.14 POTENTIAL THREATS TO CONSTRUCT VALIDITY .................................................................................. 171

TABLE 3.15 THREATS TO EXTERNAL VALIDITY .................................................................................. 168

TABLE 3.16 DEMOGRAPHIC INFORMATION OF PARTICIPANTS IN STUDY TWO ........................................................................ 169

TABLE 3.17 CHARACTERISTICS OF EXPLICIT TEACHING RESEARCH-BASED STRATEGIES .......................... 177

TABLE 4.1 PEDAGOGICAL CONTENT KNOWLEDGE PRE-POST MEAN AND STANDARD DEVIATION SCORES BY CONDITION ........................................................................ 189

TABLE 4.2 QUIZ OCCASION PEDAGOGICAL CONTENT KNOWLEDGE MEAN AND STANDARD DEVIATION SCORES BY CONDITION ........................................................................ 191

TABLE 4.3 PATTERN LANGUAGE PRE-POST MEAN AND STANDARD DEVIATION SCORES BY CONDITION 194

TABLE 4.4 QUIZ OCCASION PATTERN LANGUAGE MEAN AND STANDARD DEVIATION SCORES BY CONDITION ........................................................................ 196

TABLE 4.5 QUIZ OCCASION PATTERN LANGUAGE SOPHISTICATION MEAN AND STANDARD DEVIATION SCORES BY CONDITION ........................................................................ 199
List of Figures

FIGURE 2.1. REQUIREMENTS FOR SUCCESSFUL PRACTICE OF INCLUSION IN SCHOOLS. ........................................ 13

FIGURE 2.2. THE USE OF RESEARCH-BASED PEDAGOGIES IN SCHOOLS..................................................... 18

FIGURE 2.3. THE CAPACITY TO COMMUNICATE AND COLLABORATE ABOUT STUDENT LEARNING USING A PROFESSIONAL PATTERN LANGUAGE. ................................................................. 39

FIGURE 2.4. POSITIVE SELF-EFFICACY ABOUT INCLUSION IN SCHOOLS.................................................. 45

FIGURE 2.5. TRANSLATION OF SKILLS AND ATTITUDES FOR SUCCESSFUL INCLUSION IN SCHOOLS. .............. 48

FIGURE 2.6. THE INTEGRATION AND EVALUATION OF SKILLS AND ATTITUDES IN PTE COURSES AND THE POTENTIAL OF THEORETICAL DESIGN FOR PTE COURSES.............................................. 67

FIGURE 2.7. THEORETICAL COURSE DESIGN THAT ENHANCES INCLUSIVE SKILLS AND ATTITUDES.............. 82

FIGURE 3.1. THE APPLICATION OF EMBEDDED DESIGN...................................................................................... 104

FIGURE 3.2. STUDY ONE: NON-EQUIVALENT CONTROL GROUP DESIGN...................................................... 108

FIGURE 3.3. DATA COLLECTION PROCEDURES FOR STUDY ONE................................................................. 146

FIGURE 3.4. STUDY TWO: NON-EQUIVALENT CONTROL GROUP DESIGN.................................................... 164

FIGURE 3.5. DATA COLLECTION PROCEDURES FOR STUDY TWO.................................................................. 174

FIGURE 4.1. PEDAGOGICAL CONTENT KNOWLEDGE AT PRE- AND POST-TEST OCCASIONS BY CONDITION FOR STUDY ONE. ........................................................................................................ 190

FIGURE 4.2. QUIZ OCCASION PEDAGOGICAL CONTENT KNOWLEDGE MEAN SCORES BY CONDITION FOR STUDY ONE. ........................................................................................................ 193

FIGURE 4.3. PATTERN LANGUAGE FREQUENCY PRE-POST MEAN SCORES BY CONDITION FOR STUDY ONE. ........................................................................................................ 195

FIGURE 4.4. MEAN SCORES FOR PATTERN LANGUAGE BY CONDITION FOR STUDY ONE. ...................... 198

FIGURE 4.5. QUIZ OCCASION PATTERN LANGUAGE SOPHISTICATION MEAN SCORES BY CONDITION FOR STUDY ONE. ........................................................................................................ 200

FIGURE 4.6 SEIPD MEAN SCORES BY CONDITION FOR STUDY ONE.......................................................... 202

FIGURE 5.1. THE APPLICATION OF COURSE DESIGN TO FACTORS RELEVANT TO SUCCESSFUL INCLUSIVE PRACTICE................................................................. 219
Chapter 1: Introduction

While there is widespread support for university-based preparation of teachers for inclusion, there is also national and international concern about whether the preparation is adequate (Forlin, Kawai, & Higuchi, 2015; Husebo, 2012; Snyder, 2012). Empirical research about the efficacy of teacher education for inclusion is rarely carried out (Burkhardt & Schoenfeld, 2003; Grima-Farrell, 2012; Hempenstall, 2006; Ure, Gough & Newton, 2009; Zundans-Fraser, 2014).

This research seeks to determine the effect of theory-based teacher preparation on professional learning and self-efficacy, and the extent to which that learning translates into better inclusive practice in classroom settings. The rationale for the research focuses on the need for teacher preparation in inclusive education to make better links between theory, attitudes and beliefs, knowledge, skills and practice. The chapter employs an account of the current state of pre-service teacher education programs for inclusion as a background to and rationale for the research.

1.1 The Importance of Inclusion and Inclusive Teacher Practice in Schools

Australian and international policy requires that students with special needs receive their education in the least restrictive setting along a continuum of available placements (Florian, 2011; Jenkins, 2002; Kretlow & Helf, 2013). Many teacher graduates find themselves teaching students with varying needs within mainstream classes in inclusive schools (Ashman & Elkins, 2012; Dempsey, Foreman, & Jenkinson, 2002; Hueng, 2006). As a result, all teachers, including new graduates, are required to teach students who may vary widely in aptitude, learning history, and achievement in regular classrooms (Foreman, 2011; Mastropieri & Scruggs, 2013).

The debate about inclusion has now shifted from whether to include students with special needs to how it should take place (Alquraini & Gut, 2012; Forlin, et al., 2015; Goodman & Burton, 2010). The successful inclusion of students with different learning needs occurs when there is a confluence of teacher skill (e.g., Carter, Stephenson, & Strnadova, 2011a; Stephenson, 2003), broader school and system capacity (Bell & Dempsey, 2001; Carter, Stephenson, & Strnadova, 2011b) and a responsive curriculum (Dempsey et al., 2002; Gettinger & Stoiber, 2012; Harris, 2011).
Legislation and policy support the capacity and processes of inclusion at school and systems levels. The presence of effective capacity building features of inclusive practice, such as teacher skill and curriculum differentiation, are increasingly viewed as key benchmarks for an effective inclusive classroom that translates the legislation and policy into practice (Ashman & Elkins, 2012; Berends, Bodilly, & Nataraj, 2002). Specifically, these benchmarks required for successful inclusion include:

- Teacher use of research-based pedagogies required to enhance success with students in schools (Dean, Stone, Hubbell, & Pittler, 2012; Evmenova & Behrmann, 2011; Hattie, 2009; Kretlow, Cooke, & Wood, 2012; Kyndt et al., 2013; Magnesio & Davis, 2010; Rosenshine, 2012);
- Teachers’ capacity to communicate in a sophisticated manner about student learning (Allen & Blackston, 2003; Crafton & Kaiser, 2011; Friend & Cook, 2013; Hennessey & Deonigi, 2013; Knowlton, 2004; Levin, 1999; Schaubman, Stetson, & Plog, 2011);
- Teachers’ sense of self-efficacy when working with students who have a variety of learning needs (Abbitt, 2011; Al-Awidi & Alghazo, 2012; Avery & Meyer, 2012; Chiner & Cardona, 2013; Hastings, 2012; Holzberger, Philip, & Kunter, 2013; Jamil, 2012; Yilmaz, 2011); and

Research-based pedagogies including explicit teaching (ET) and cooperative learning (CL) were chosen as the vehicle to illustrate this study’s theory to practice intent because they possess a convincing evidence base including statistically robust effects in controlled experimental research over time, settings and populations (e.g., Bloom, 1984; Fraser, Walberg, Welch, & Hattie, 1987; Gillies & Ashman, 2000; Hattie, 2012; Slavin, 1991). Despite strong research support, ET and CL have not been employed as routine practice in many schools or inclusive classrooms (Klingner, Vaughn, Hughes, & Arguelles, 1999; Vaughn, Linan-Thompson, & Hickman, 2003). Carter et al. (2011b) confirmed that models of ideal practice found in the research literature do not line up with actual practice in schools. Studies have found teacher understanding of research-based pedagogies to be inconsistent and implementation of them in practice to

Sophisticated levels of professional language are required for communities of teachers to work together to discuss these pedagogies and problem solve in the complex inclusive learning environments they encounter (Ainscow, Booth, & Dyson, 2004; Ainscow & César, 2006; Burke, 2002; Cochran-Smith, 2005; Hargreaves, 1994; Ware, 2003). Even though insufficient time is frequently given as a rationale for the lack of collaborative processes in schools, studies have also indicated that there is a lack of teacher expertise with the professional language associated with the evidence-based practice required to collaborate with colleagues (Crafton & Kaiser, 2011; Schaubman et al., 2011). When these capacity building skills of consistent professional language and the use of pedagogical expertise are evident, it is expected that self-efficacy and competence of graduate teachers for working with students who have varying needs would also be increased (Chan, 2008; Hastings, 2012; Pajares, 1992). Currently, teacher self-efficacy is shown to be alterable when such skills are mastered (Chan, 2008; Jamil, 2012).

Despite the existence of legislation and policy, the aforementioned benchmarks for the inclusion of students with special needs remains a challenge for teachers and especially the translation of knowledge about research-based pedagogies into routine practice (Bauer, Johnson, & Sapona, 2004; Foreman, 2011; Mastropieri & Scruggs, 2007; Wolfe & Hall, 2003).

The following section will outline the role of pre-service teacher education programs in preparing pre-service teachers for the challenge of inclusion.

1.2 The Importance of Inclusion and Inclusive Teacher Practice in Pre-service Teacher Education Programs

Highly qualified, skilled teachers are essential to carry out efficacious research-based interventions and solve problems that arise in complex learning environments (King-Sears, Boudah, Goodwin, Raskind, & Swanson, 2004). Carroll, Forlin, and Jobling (2003) suggest that programs for preparing pre-service teachers are not adequate in terms of developing these skills and closing the gap between what is known about inclusive practice and what happens in inclusive classrooms. According to Carroll et
al. (2003) and Dempsey et al. (2012b), teacher preparation programs frequently over-emphasise knowledge acquisition to the detriment of equipping teachers with practical skills for teaching a diverse range of students with varying abilities. In the Australian context, the Teacher Education Ministerial Advisory Group (TEMAG) noted that although there is a growing international focus on research and evaluation of initial teacher education programs there is a lack of Australian research about what practices and approaches have an impact. The group call for a rigorous evidence base about effective initial teacher education that would allow programs to learn from and be benchmarked against effective practice (AITSL, 2015).

The design and composition of pre-service programs and their impact on future inclusive practice is particularly important in this regard (Billingsley, 2004; Billingsley, Carlsen, & Klein, 2004; Darling-Hammond, 2005; Darling-Hammond, Hammerness, Grossman, Rust, & Shulman, 2005; Fore, Martin, & Bender, 2002). One of the recommendations from the Australian Teacher Education Ministerial Advisory Group states that in order for universities to be granted full program accreditation, “…they will need to clearly show that their graduates are classroom ready, how their graduates are having a positive impact on student learning, and that employers are satisfied with the graduates they produce.” (TEMAG, 2015, p. 5). The literature related to pre-service teacher education courses identified a number of factors that have contributed to the adequacy or inadequacy of PSTs’ preparation for addressing the challenges associated with inclusion. Many of these overlap with the factors that had been identified in schools and include: insufficient depth of preparation (Forlin, Tait, Carroll, & Jobling, 1999); inadequate linkage of the classroom to professional experiences (Koehnecke, 2000; Mitchell, Schwager, & Doolittle, 1997); empirical evaluation of these factors; and the lack of a common theory base driving the design coherence of teacher education programs (Francis, 2002; Mitchell et al., 1997; Roth, Lawless, & Tobin, 2000; VanderVen, 2000). The need for more responsive and coherent teacher education programs has been consistently documented in the broader education literature (e.g., Darling-Hammond, 2006; Fullan, 1999; TEMAG, 2015).

Currently there are concerns about the limited research undertaken regarding the effectiveness of teacher education (Sleeter, 2014). A key problem seems to be that many teacher education programs keep repeating the same practices they criticise (Zundans-Fraser, 2014). For example, although many respondents to the *Quality
Matters: Inquiry into Teacher Education (Ramsey, 2000) in Australia noted that evidence-based research was promoted as critical in the field and as a key way to identify and validate effective models of teacher education, little of this style of research was actually carried out (Burkhardt & Schoenfeld, 2003; Hempenstall, 2006; Ure et al. 2009). Fifteen years after the Ramsey review, improving teacher education remains a key component of the Australian Government’s Students First policy. Having concerns about the limited research occurring in teacher education, the Australian Government convened a panel of experts to deliver an urgent report comprising 38 recommendations aimed at ensuring graduates are better prepared for the future challenges they will face in the classroom. A new Higher Education Standards Framework legislative instrument was made by the Minister for Education and Training, Senator the Hon Simon Birmingham, in October 2015 which outlines the basis for the regulation of higher education providers and courses by the Tertiary Education Quality and Standards Agency (TEQSA). The Standards in this framework came into effect from 1 January 2017.

Based on these concerns about the limited evaluation of inclusive education courses, it is imperative that more empirical evaluation is carried out.

1.3 Evaluating the Courses Used by Faculties to Prepare Pre-service Teachers for Inclusion

Throughout the thesis the term ‘course’ is used to refer to a single unit (subject) of study for PSTs within their Pre-service teacher education program. Empirically evaluating courses (single units of study) used by faculties in higher education PST programs remains of highest importance (Dawkins, Dickerson, McKinney, & Butler, 2008; Forlin, 2010a; Szabo, Scott, & Yellin, 2002). There is a need to investigate whether teacher education for inclusion works, whether theoretical underpinnings are used to teach research-based pedagogies; and to investigate the links of that pedagogical knowledge to professional experience. Cochran-Smith and Zeichner (2005) conducted a major study in the United States within teacher education and developed a substantial evidence base for practices found in pre-service teacher education. The study noted five major elements required when preparing teachers for successful inclusion: (a) establishing a shared language in teacher education curriculum to support collaboration between special and general educators; (b) establishing collaborative clinical practices; (c) ensuring competence in pre-service teachers before graduation; (d) supporting beginning teachers in the first three years
after graduating; and (e) promoting shared governance that reflects collective responsibility between stakeholders in pre-service teacher education. At the time of Cochran-Smith’s (2005) review, she noted that studies of pre-service teacher preparation specifically focused on working with students who have additional needs were sparse with only 17 empirical studies identified.

Given the lack of empirical investigation into teacher preparation for practice, the inclusion of research-based pedagogies in PST programs that are known to have the greatest effect on school student learning need to be interrogated more rigorously. When examining the practices used in pre-service teacher education, the majority of studies employed participant-completed survey-based data (Hamre & Oyler, 2004; Hemmings & Woodcock, 2011; Mitchell et al., 1997), reflective essays, and self-study, or self-reflection (Major, 2011) as outcome measures. While these can be rich sources of data, it is important to also develop a more objective empirical understanding of the constellation of factors that impact on PST’s sustained use of evidence-based effective pedagogies in inclusive classrooms. Other researchers have used observational data to investigate the integrity and efficacy of pedagogy implementation (Huberman & Miles, 1984; in Gersten, Baker, & Lloyd, 2000) or focused on understanding how PSTs are impacted by attitudes (Carroll et al., 2003; Shade & Stewart, 2001). The extant body of work on teacher education programs is useful to the extent that it provides measures of PSTs attitudes towards inclusion. However, attitudes alone constitute a limited definition of what enhances teacher effectiveness in their practice.

Only in instances where PST outcome data are actually measured can the impact of teaching practices and pedagogies taught in pre-service teacher education courses be empirically validated. The present study uses empirical investigation to determine if a PST course is teaching participants those skills and attitudes that will assist with their inclusive classroom practice.

To date, no evidence of a validated process has been located that showed the effect of PST preparations on correlates of requisite skills (i.e., pedagogical content knowledge, professional language about research based pedagogies, self-efficacy) and actual PST behaviour in inclusive classrooms. Theoretical frameworks are required in pre-service teacher education programs to provide a template for the maintenance and sustainability of these skills acquired. Empirically validated theoretical frameworks
would enable consistency of reliable practices.

1.4 Theoretical Frameworks Evident in Pre-service Teacher Education Programs

According to Francis (2002), Roth (1999) and VanderVen (2000), theory should be advocated as a means to drive the practices to improve pre-service teacher outcomes. The extent to which the field of PST education is driven by theory is not clear from the literature currently available.

Many authors (see for example: Bates, 2002; Drever & Cope, 1999; Galguera, 2011; Kunzman, 2003), suggest the field of teacher preparation is devoid of coherent theoretical design bases. In addition, there is little evidence in the extant literature that those theories espoused are validated in actual practice. For example: conceptual frameworks are discussed (Bauer et al., 2004; Sindelar, 1995); however, few are evaluated empirically (Gersten & Smith-Jones, 2001; Mitchell et al., 1997). Cochrane-Smith suggested that one of the most prominent issues in teacher education identified in the literature both nationally and internationally was the call for uniform standards based on empirical evaluation and consistency across preparation and accreditation programs (Cochran-Smith, 2004b, p. 297; House of Representatives & Standing Committee on Education and Vocational Training, 2007), as well as the simultaneous agenda of improving teacher education through professionalisation (Cochran-Smith, 2004a; Evans, 2010; Phillips, 2008).

Researchers agree about the need for a better quality pre-service program structure; however, few have been able to articulate how to go about developing such an approach with reference to an empirically validated theory. There is a difference between coming up with models that meet accreditation requirements, and whether they are theorised and empirically defensible based upon rigorous evaluation.

There are examples in the literature related to the practice of in-service teachers and related to school reform that can be educative in terms of reform, but there are no comparable pre-service studies. Little and Houston (2003) were able to demonstrate an empirical process in the school settings by providing a conceptual theoretical framework as well as specific implementation activities to ensure best practices were used in school settings. The professional development process proposed included four
steps, each of which encompassed observable criteria that could be used to judge their reliability, validity, and relationship to the theoretical framework.

Just as Little and Houston (2003) were able to provide specifics of an empirical process of implementation in school settings, Bain (2007) was able to demonstrate an empirical process of implementing a theoretically driven school reform. In response to the need described in this literature, Bain (2007) posited a meta-theory grounded in complexity (Brodnick & Kraft, 1997; Davis & Sumara, 2006; Prigogene & Stengers, 1984) and self-organising systems theory (Wright, 2000), to enact research-based practices and to ensure the school teachers employing them were capable of dealing with changes in environmental issues, and challenges, in ongoing and sustainable ways.

Self-organising systems theory has been applied across a variety of human systems. It has been used across fields of health care as a means of delivering improved interventions of primary care in complex health care systems (Litaker, Tomolo, Liberatore, Stange, & Aron, 2006); and in instances across the complex arenas of business (Berreby, 1996; Houchin & MacLean, 2005; Tsoukas & Hatch, 2006). In education, its potential has been discussed by Mason (2008), and applied to learning (Davis & Sumara, 2006), the design and function of secondary schools (Bain, 2007), and school leadership (Walker & Dimmock, 2000). Cochran-Smith et al. (2014) suggests that to date, studies of teacher education have tended to be

“…focused on pieces of teacher education rather than wholes, and have used an underlying linear logic. It may be, however, that what is needed are new research questions and theoretical frameworks that account for wholes, not just parts, and take complex, rather than reductionist perspectives” (p. 1).

They concur that complexity theory, from which self-organising systems theory has evolved, may offer a suitable framework in which to understand the complex ‘whole’ of teacher education (Cochran-Smith, Ell, Ludlow, Grudnoff, & Aitken, 2014). Further elaboration of complexity theory is included in section 2.6.3.

The principles from self-organizing systems applied in this study were derived from a design theory by Bain (2007) based upon the application of the principles of complexity and self-organization to education. The design theory and approach known as self-organizing schools (SOS) were tested in a validation study by Bain (2007) conducted over a
decade in a school setting with students who have diverse learning needs. The focus of that work was to employ a practical theory of school design to embed research-based practice in the life of a school. Among other findings, this work reported positive relationships between the application of the theoretical principles and the increased use by teachers of research-based pedagogies (Bain, 2007). This work showed that teachers in an inclusive school could become more productive, more collaborative, and responsive to learner needs when that school was designed using principles of self-organising systems (Bain, 2007). In 2011, Bain et al. positioned a high school as a self-organising system to demonstrate the ways that the same principles enhanced the capacity of staff and students involved (Bain, Walker, & Chan, 2011). The aforementioned outcomes for teachers pursued by Bain (2007) are consistent with the challenges faced by pre-service teacher educators in the need to make their programs for pre-service teachers more responsive to the demands of the inclusive classroom. The Australian Institute for Teaching Standards and Leadership recommend that any graduates from a Higher Education Institute demonstrate they can meet the Standards and provide evidence of their impact (AITSL, 2015).

When considering the absence of theory and the lack of validation of courses at the pre-service level, it makes sense to test whether these in-service gains could be translated to the pre-service level. As a consequence, a program of research was initiated at a regional Australian university which produced a number of preliminary studies applying one of the principles from the SOS approach in school reform to the pre-service setting (Bain, 2007). This research used a case study approach to investigate the impact of the key features of the SOS on the dependent measures such as pedagogical content knowledge (Bain, Lancaster, Zundans, & Parkes, 2009) and pattern language (Bain, Lancaster, & Zundans, 2009). In the earlier studies, these important inclusive dependent measures were investigated employing a within subject design and were used as a basis that informed this thesis. The current study expanded on these results and investigated the variables using a between subject design and included an investigation to determine if the theoretical principle of embedded design, would exert a differential effect on the actual practice of pre-service teachers when they were in the field.

Complexity and self-organising systems theories have created a useful framework for this study because pre-service education courses are dynamic, sophisticated and deal
with complex interactions at multiple levels and contexts. SOS, as applied to educational contexts, represents an emergent natural (or logical) order that can be used to explain the dynamic way the elements of such complex systems interact. The work of Bain (2007) and subsequent studies (Bain, Lancaster, & Zundans, 2009; Bain, Lancaster, Zundans, & Parkes, 2009) found that the complex systems operating in schools could work successfully when the characteristics and principles of SOS are applied and nurtured within their contexts context.

This study was situated within this framework of complexity and self-organising systems theories. Specifically, the study focussed on applying the principle of embedded design derived from (SOS) theory to a pre-service teaching program and in particular, to the preparation of undergraduate teachers the practice for inclusion. The embedded design principle is the key theoretical principle of the SOS approach as it entails embedding essential characteristics throughout all practice when designing an inclusive education course.

1.5 Summary
In conclusion, there is a need to ensure that pre-service teachers possess the skills required for inclusive practice and the capacity to communicate that practice in professional ways as they collaborate with others. The current body of literature has not demonstrated the efficacy of pre-service teacher education in providing graduate teachers with those skills nor whether pre-service teacher courses are teaching the skills to requisite levels of mastery that can be maintained and generalised to inclusive contexts.
That is, the literature demonstrates that practicing teachers do not implement these inclusive practices with fidelity.

Guidance exists that indicates what inclusive education courses require to ensure their graduates have the requisite skills, knowledge and attitudes to work effectively in the field. There is also support for the role of theory to design the courses, but efficacy studies are sadly lacking.

While theory can provide a focus and organising framework for the design of Pre-service teacher education, the literature to connect theory to the design of courses is limited. This lack of theory inhibits efforts to determine the efficacy of PST
preparation for inclusive education given the lack of an anchoring framework or set of
design principles that can drive efficacy research.

The studies undertaken in this thesis stand at the confluence of these issues. The
research seeks to determine whether the implementation of a theoretically derived
course design covaries with the development of pedagogical knowledge, self-efficacy
and pattern language of pre-service teachers, and whether those competencies are
expressed in actual classroom practice. As such, the study seeks to contribute a more
robust base of evidence concerning teacher preparation for inclusion.

Specifically, the following gaps within the literature have been identified as priorities
for further exploration and analysis. These variables are integrated throughout the
current research study:

a) Developing PST proficiency and mastery of research-based pedagogies that
may be translated into practice in inclusive classrooms and settings;
b) Incorporating professional language to support collaborative practices in
complex learning environments of inclusive education;
c) Improving PST self-efficacy for teaching students with varying needs;
d) Addressing the inadequate theory base currently evident in PST programs;
and
e) Embedding empirically tested theoretical frameworks in the design of
teacher education programs which is thought to allow cohesion of the
above elements.

Chapter 2 expands the rationale described here by reviewing the literature that informs
this research and the formulation of the research questions. Chapter 3 describes the
methods used to answer the research questions. Chapter 4 presents the results obtained
from the data collected and analysed. Chapter 5 unpacks and interprets the results from
this study in relation to other findings from the literature; suggests limitations within
the current study; and finally suggests directions for future research.
Chapter 2: Review of the Literature

Discussions in the literature surrounding teacher preparation for inclusion are extensive (Jung, 2007; Killoran, Woronko, & Zaretsky, 2014; Romi & Leyser, 2006), yet there are few empirical studies specifically focused on the way teacher preparation courses influence pre-service teachers’ learning about and practice of inclusion.

This review of the literature covers two main sections. Following a brief introduction to inclusion, the factors that are shown to be required for successful inclusion in schools are described and discussed. The range of factors that relate to the efficacy of inclusive practice include: the use of evidence-based pedagogy; teachers’ capacity to communicate about pedagogies and the professional language they use to do so; and their attitudes, beliefs and dispositions about inclusive practice. Then the areas of pre-service teacher preparation and the ways the indicators of successful inclusion are developed and evaluated in pre-service teacher education programs are reviewed. The review integrates the larger commentary literature with the smaller number of research studies that have examined the impact of course design on the disposition and professional learning of PSTs.

The process used to conduct the review of the literature about inclusive practice included locating applied research published in peer reviewed journals since 1990. This timeframe was selected as this is when educational context, including curriculum differentiation and use of appropriate teaching pedagogies finally emerged as a focus for meeting student learning needs rather than focusing exclusively on the deficits of the child (Ashman & Elkins, 2012; Foreman, 2011). Up until this time there had also been extensive efficacy research on pedagogies used with students who have additional learning needs. Computer searches were conducted using both the PsychINFO and ERIC databases. To assist in locating relevant studies the following key words were used: successful inclusion; special education; inclusive teacher practice; research-based pedagogies; attitudes; self-efficacy; capacity for communication; and collaboration. In addition to the computer searches, reference lists from syntheses of teacher practice and research were examined to provide additional sources of information.
The discussion that follows is organized according to the findings of the review. The literature indicates that the following are required for successful inclusion:

- Inclusive teacher practice in schools: What is it? The importance of quality practice. What do teachers think about inclusion and why does it remain a challenge?
- Teacher use of research-based pedagogies to enhance inclusion;
- Teachers’ capacity to communicate about student learning using professional pattern language at a sophisticated level;
- Teachers’ self-efficacy for success in the inclusive classroom; and
- Translation of research-based pedagogies, skills and attitudes into classroom practice.

The aforementioned factors relating to the success of inclusion are represented in Figure 2.1 where the funnel represents the findings and the organization of the review diagrammatically. The intentional use of a funnel is to indicate all requirements that are needed by teachers to enable successful inclusion in practice. This figure is used to highlight each of the sections as they are discussed throughout the chapter.

![Figure 2.1. Requirements for successful practice of inclusion in schools.](image)

2.1 Inclusive Teacher Practice Context

2.1.1 A brief introduction about the nature of inclusion.
Inclusion is the process of increasing participation and decreasing exclusion from the culture, curriculum and community of mainstream schools (Florian, 2011). Australian and international legislation requires that students with special needs are educated in
the least restrictive setting along a continuum of available placements (Jenkins, 2002). As a result, all teachers, including new graduates, are required to teach students who may vary widely in aptitude, learning history and achievement in regular classroom settings (Ashman & Elkins, 2012; Hueng, 2006). The successful inclusion of students with different learning needs occurs when there is a confluence of teacher skill, broader school and system capacity, and a responsive curriculum (Foreman, 2011; Gettinger & Stoiber, 2012; Harris, 2011). Effective knowledge, skills and attitudes of practitioners are required to help bridge the gap between guidelines, policy and actual implementation of inclusive practices.

Legislation and policy are designed to support the capacity for and processes of inclusion at school and systems levels. Australian federal disability discrimination law, the Disability Discrimination Act 1992, along with the Federal Disability Standards for Education 2005, enact specific requirements for education (Cumming & Dickson, 2013). In Australia, the Commonwealth legislated Disability Discrimination Act (DDA) in 1992 was intended to increase awareness of students who have a disability and their inherent rights to have their needs met appropriately rather than being excluded from educational activities (Cumming & Dickson, 2013).

For teachers supporting students who have additional needs and are included in regular classrooms, the relationship between school and system capacity and teacher skill is explained in the following way: Our practice will be influenced by our attitudes and beliefs (principles), by laws (legislation), and by organisational guidelines (policy). It is an interactive process that needs to evolve and reflect changes in society as they occur over time (Foreman, 2011 Hueng, 2006).

2.1.2 The history leading to current practices.
The history and genesis of inclusive practice has taken several notable turns over the past half century (Kavale & Forness, 2000). Three major factors of access, categorisation, and context have shaped the current forms of inclusive education that we see in operation today. Factors involving access to education emerged in the 1970s and resulted in legislation that allowed students access to the education system where they had previously been denied. Early efforts for service provision focused on diagnosis and categorisation of handicapping conditions as a means of determining the most suitable type of service for a student. Now, context, including curriculum
differentiation and use of appropriate teaching pedagogies have emerged as the focus on features of the actual learning environments instead of exclusively on the child (Ashman & Elkins, 2012; Foreman, 2011). Inclusion then occurred as an indicator of place along a context driven continuum, with inclusion in the regular class as the most preferred end of the continuum (Jenkins, 2002).

Rather than inclusion being seen as a singular place along the continuum, it is now regarded as a process to be followed and determined by all impacting contextual factors (Jenkins, 2002). Inclusion is supported by the belief that all children, regardless of ability, belong together in one classroom (Mastropieri & Scruggs, 2013). The inclusive school puts students at the centre of all instructional decisions (Kugelmass, 2001). Thus, over time the debate about inclusion has evolved from a conversation about whether or not to include students in the mainstream classroom, to a focus on how it should take place (Alquraini & Gut, 2012; Forlin, et al., 2015; Goodman & Burton, 2010).

The passage of legal mandates ended the debate about whether inclusion should take place. In Australia, Commonwealth legislation (Disability Discrimination Act, 1992) was implemented to increase awareness of people who have a disability and their inherent rights to be treated as everyone else and not excluded. The Federal Government impacts policy and practices of the States through the provision of discretionary funding (Cumming & Dickson, 2013). However, there still remains wide variation between States in the emphasis given to the form a provision will take (i.e., from enrolments offered between regular, special class and special school placement). While the Disability Discrimination Act of 1992 was designed to have an impact on ‘awareness’, it did little to direct the actual classroom practices of teachers (Dempsey, 2004). The Disability Standards for Education were developed in 2003 and ratified in 2005 in an attempt to address this issue (Cumming & Dickson, 2013). These Standards now specify enrolment, participation, curriculum and student support services to be adhered to by schools (Dempsey et al., 2002). As a result, the ratification of these Standards supplements the legislation and provides clarity around what is involved to address issues such as segregation and lack of differentiation.

In order to articulate how inclusion takes place, a clearer definition of the inclusion construct is needed to operationalise it in practice, describing it as a ‘process’ and not
as a state or a setting (Darlington, 2003). Fox and Ysseldyke (1997) refer to inclusion as a practice that operationalises the concept of the “least restrictive environment” as well as other social justice principles such as “all children can learn”. Instruction should be individualised. The local school is one option for enrolment, and it is the responsibility of class teachers to meet the needs of all children in their classes (Bell & Dempsey, 2001; Dempsey et al., 2002).

2.1.3 The importance of quality inclusive teacher practice.

Inclusion requires that teachers take responsibility for all students in their classes, however, not all feel they have the confidence to do this well. In a synthesis of 28 surveys by Mastropieri and Scruggs covering the years 1958-1995 (1996, cited in Mastropieri & Scruggs, 2013), 30% of teachers thought they had enough expertise and training for inclusion to be successful; the remaining 70% did not share this level of confidence. This could be seen as understandable apprehension when the field has indicated for over 20 years prior that a separate system is appropriate and better suited to the needs of students with disabilities. Nolte (2010) and Richards (2010) suggest effective inclusion can only be achieved when those involved are able to participate confidently.

Wolfe and Hall (2003) indicate that effective inclusion can be achieved through teacher capacity building on how to bring research-based interventions to the classroom and how to individualise programs appropriately. In inclusive settings, the Individual Education Program (IEP) is the basis for high-level intervention as described above. The role of the IEP is to determine what should be taught, how the content should be taught, and who can most appropriately provide the instruction (Wolfe & Hall, 2003). Ysseldyke (1999) identifies these same essential elements as necessary requirements to improve teaching and learning more generally. A teacher needs to know the goals the students are striving for, how to achieve these goals (i.e., what pedagogies to use); and, finally, if and when the goals have been achieved. These elements are also reflected in the principles, legislation and policies of inclusion discussed above. For many children in Australia, IEPs are still not being provided (Dempsey, 2012b).

With prevalence rates of learning difficulty reported in the literature varying upwards of 43% (McKinnon & Gordon, 1999; McLeod & McKinnon, 2006; Mitchell,
Schwager, & Doolittle, 1997), ensuring that all teachers can address variability in student learning needs is a critical issue. An attempt was made to address this need with the introduction of mandatory pre-service teacher education courses in inclusion across all Australian university programs in 1993 to place mastery of these strategies within the outcomes expected for PSTs completing their teaching program. However, there is a view that these courses have not adequately addressed the lack of teacher skill (Forlin, 2010a, 2010b; Lastrapes & Negishi, 2012; Romi & Leyser, 2006). In NSW, the current expectations of newly graduated teachers are that teacher skill would be at the beginning of the continuum of a “Proficient Teacher” as per AITSL Standards 3.1 – 3.7 (AITSL, 2015).

2.1.4 What do teachers think about inclusion and why does it remain a challenge?
Despite requirements based on federal legislation related to inclusion, there remain significant challenges to inclusive practice and there is little evidence of change in the way schools operate (Hueng, 2006). The legislation-policy-research-practice gap remains an enduring problem in the field of education (Grima-Farrell, 2012). Researchers in the field of inclusion concur there is a need for high-fidelity implementation of research-based teaching pedagogies to provide a unified system of education where all students might potentially be educated in mainstream settings (Kretlow & Helf, 2013; Mastropieri & Scruggs, 2013; Rosenshine, 2012).

2.1.5 What do teachers need to know to make inclusion happen?
This section focusses on pedagogies to promote successful inclusion. Responsiveness to learner diversity is increasingly viewed as a benchmark of teacher effectiveness in inclusive classroom settings (Ashman & Elkins, 2012). Reaching this benchmark requires that teachers successfully deploy well-researched teaching and collaborative approaches to classroom instruction (Mastropieri & Scruggs, 2013; Rosenshine, 2012). Teacher use of research-based pedagogies and curriculum differentiation with high-quality pedagogies are required to enhance success for students in schools. Kretlow and Helf (2013, p. 168) suggest that “when evidence-based … instruction is implemented with fidelity; instruction can be eliminated as a reason for students not making adequate progress”. A sound understanding of pedagogical content knowledge about research-based strategies will assist teachers in translating these into practice. Figure 2.2 positions pedagogies as one of the variables required to translate inclusion into successful practice.
2.2 Teacher Use of Research-Based Pedagogies Required to Enhance Inclusion

This section takes up discussion of explicit teaching and cooperative learning as two of the most highly researched pedagogies with the strongest efficacy available to teachers in inclusive classrooms. These are also the approaches used as a focus in this study. Explicit Teaching (ET) and Cooperative Learning (CL) are consistently identified as essential pedagogies for successful teaching in inclusive classrooms (Rosenshine, 2012). The skills required by teachers for differentiating classroom instruction include, among other things, advanced knowledge of these evidence-based instructional approaches (Mastropieri & Scruggs, 2013).

CL and ET possess an extensive research history that shows statistically robust effects in controlled experimental research over time, settings and populations (e.g., Fuchs & Fuchs, 1998; Gillies & Ashman, 2000; Hattie, 2009, 2012; Slavin, 2005). The effectiveness of these practices makes them widely advocated and accepted features of programs that enable teachers to work successfully in inclusive settings. As exemplars of research-based pedagogies, ET and CL are used in this thesis as the vehicles to illustrate the ways the theory to practice intent is made possible. The first pedagogy of focus was explicit teaching. Following is a definition of that pedagogy and rationale for its use, research support, and issues and challenges teachers face when using the pedagogy.
2.2.1 Explicit teaching.

2.2.1.1 Explicit teaching description.

Explicit teaching (ET) is a pedagogy that enables the conditions for curriculum and instruction to be differentiated. It involves systematic teacher demonstration, modelling, guided practice and the provision of feedback. It is useful for all students, including those who have complex support needs and require many adaptations to curriculum and instruction to enable learning to occur (Darch, Eaves, Crowe, Simmons, & Conniff, 2006; Killen, 2007; Rosenshine, 1986, 1987; Ryder, Burton, & Silberg, 2006). ET is a pedagogy that promotes considerable teacher direction. It is most commonly used when a new area of study is being introduced. A broad overview of ET is now given with key concepts defined to assist with foundational knowledge. ET usually involves the following characteristics:

- Learning outcomes and success criteria are made clear to students;
- The teacher directs the lesson activities and time;
- Small steps are devised and successful student practice is required at each step;
- Feedback is given until independent mastery is reached; and
- Students are carefully monitored throughout the process. (Pace, 2011, p. 7)

Hattie (2009) describes the underlying principles of direct instruction as having major characteristics which match the above steps. In ET, the teacher decides the learning intention and the success criterion, makes the intentions transparent to the student, demonstrates the requirements by modelling, provides guided practice until mastery is reached, and evaluates if understanding has occurred during independent practice (Hattie, 2009, pp. 205-206). In this thesis, the term explicit teaching is used but it incorporates all the characteristics described by Hattie as direct instruction.

2.2.1.2 Meta analyses that provide a rationale for the use of explicit teaching.

The most compelling rationale for the use of ET is found in the results of meta-analyses that have been carried out over time on the features that comprise explicit teaching. Hattie (2009) presented a synthesis of over 800 published meta-analyses relating to strategies that have varying effects on the achievement of school-aged students. In Hattie’s analysis of teaching approaches, he noted that setting goals and having success criteria framed the challenge and purpose of the lesson and fostered student involvement. Goals are effective when they are challenging and specific; a
direct linear relationship is noted between goal difficulty and performance (with an effect size of $d = 0.67$) (Hattie, 2009, p. 164). Challenging and specific goals are more effective on student progress than vague suggestions such as “do your best” ($d = 0.66$) (pp. 164-165). Fuchs and Fuchs (1998) investigated the differences between long and short term goals and noted that short term goals are required for surface learning outcomes and long term goals are required for deep learning outcomes. Behavioural objectives were required to bridge the gap between the learner’s previous knowledge and new learning to occur. They incorporated a specific learning goal and learning success criteria to indicate when the learning target had been reached.

In statistics, an effect size is a quantitative measure of the strength of a phenomenon. Examples of effect sizes are the correlation between two variables, the regression coefficient, the mean difference, or even the risk with which something happens (Bronkhorst, Meijer, Koster, Akkerman, & Vermunt, 2013). For each type of effect-size, a larger absolute value always indicates a stronger effect. Effect sizes complement statistical hypothesis testing, and play an important role in statistical power analyses, sample size planning, and in meta-analyses (Burns, 2000). The following are accepted as small effect size (0.12), medium (0.50) and large (0.80) effect sizes when paired quantitative data are available.

Worked examples are a way of showing students what success looks like, including steps toward the required solution. This links closely to modelling steps in ET. The cognitive load is reduced and students can concentrate on the process of working toward the correct answer (Mestre & Ross, 2011). The phases of worked examples provided by Hattie (2009) include the introductory phase (exposure to the example), acquisition or training phase, and a test phase (assessment of the learning). These related specifically to the phases of ET from introduction, to modelling, and finally assessment. Overall, the effect of using worked examples as a part of instruction was medium at $d = 0.57$ (Hattie, 2009, pp. 172-173).

Feedback had the most powerful influence on achievement with large reported effect size, $d = 0.73$ (p. 173). Mastery learning, a form of explicit teaching, emphasises the use of feedback and claims that all children can learn when provided with appropriate scaffolding, feedback and time to master the material provided (Hattie, 2009, p. 170). Numerous feedback loops are required within the ET sequence, with the most powerful
being from the student to the teacher when seeking to find out what students know and understand. This feedback loop makes learning visible for both student and teacher. Most effective feedback provides cues and reinforcement to the learner. It is not about providing rewards but rather about providing information about the task completion. When embedding the feedback characteristics of explicit teaching into peer tutoring arrangements, Fuchs and Fuchs (1998) found increased achievement for students with a learning difficulty by $d = 0.70$. Teachers used evidence-based practices that included behavioural objectives, noting specific goals and achievement required, modelling and guided practice through worked examples, as well as feedback. Explicit teaching creates ample feedback opportunities provided by the teacher during the guided practice stage.

Using 134 studies and 4,000 effect sizes from various teaching methods, Marzano (1998) also found the overall effect size for instruction with the characteristics of explicit teaching was 0.65. Marzano (1998) concluded that the effective teacher was one who used clear instructional goals and provided much and deliberate practice, modelling and feedback for students to reach the goals set. These findings were subsequently supported by later studies (Haystead & Marzano, 2009; Marzano, Pickering, & Pollock, 2001).

Brophy (1986) extensively reviewed the research on teacher effectiveness and linked particular teacher behaviour to student achievement. The following were noted and are all evident in explicit teaching formats: academic learning is influenced by the amount of time spent in academic engagement; learning is easier when it is structured and related to prior knowledge; and corrective feedback should be provided. According to Killen (2007), the aspects of explicit teaching that consistently receive research support as important indicators of effective teaching include: clarity (as in cognitive clarity of task as well as verbal clarity in the way the lesson is delivered); variety (task orientation, task engagement, and high success rate); ways of checking students’ understanding; access to information; and types of feedback given.

Rosenshine (1986) concurred and described six instructional activities that remain essential to well-structured teaching: regular review of past learning; well-organised presentation (present material in small steps); guided practice for each step; ask
questions to check understanding and obtain responses from all students; feedback
given; and independent practice once the basics have been mastered.

The results of the meta-analyses from Hattie (2009), Marzano (1998), and studies by
other researchers such as Fuchs and Fuchs (1998), Brophy (1986), and others cited
here, confirm those characteristics of explicit teaching that enable the pedagogy to
achieve the most successful student outcomes. Features such as modelling, guided
feedback, and independent practice are effective for all students and especially those
with additional learning needs.

2.2.1.3 Specific research studies that lend support for explicit teaching.
Research conducted over the past 20 years lends extensive support to the use of explicit
teaching in structured as well as far more complex and less defined curriculum areas
(Hattie, 2009; Ryder et al., 2006). For example: in the area of reading instruction
(Adams & Slocum, 2004; Benner, Kinder, Beaudoin, & Stein, 2005; Cooke, Gibbs,
Campbell, & Shalvis, 2004; Fielding-Barnsley, 1997; Marchand-Martella, Martella,
Kolts, Mitchell, & Mitchell, 2006; Przychodzin-Havis et al., 2005; Pullen, Lane, Lloyd,
Nowak, & Ryals, 2005; Tobin, 2004; Van, 2004); vocabulary instruction for young
children (Spencer, Goldstein, & Kaminski, 2012); mathematics (Hofmeister, 2004;
McKenzie, Marchand-Martella, Moore, & Martella, 2004; Parsons, Marchand-
Martella, Waldron-Soler, Martella, & Lignugaris-Kraft, 2004; Stein, Kinder, &
Milchick, 2004); social skills (Close & Kreitzer, 1998); spelling (Darch et al., 2006);
teacher education and attitudes (Kinder, Kubina, & Marchand-Martella, 2005; Snider &
Schumitsch, 2006); science (Gess-Newsome, 2002; Matkins, Bell, Irving, & McNall,
2002); and writing (Martella & Waldron-Soler, 2005; Walker, Shippen, Alberto,
Houchins, & Cihak, 2006).

These results investigated the effect of using explicit teaching to teach a variety of
curricula areas to various groups of students. Hattie (2009, p. 129) investigated
curricular effects on student achievement. He summarised 50 meta-analyses on reading
and found an average effect of \( d = 0.51 \) in favour of ET over other strategies. Of the 4
meta-analyses of 64 studies on ‘whole language’ the effect size was minimal \( d = 0.06 \)
(p. 137), compared to 7 meta-analyses of 301 studies on ‘vocabulary programs’ where
the effect size was high \( d = 0.67 \) (p. 131). Phonics instruction had an average effect size
of \( d = 0.60 \) (p. 133), writing \( d = 0.44 \) (p. 142), science \( d = 0.40 \) (p. 147), and maths \( d = \)
0.45 (p. 144). In spite of the considerable effect sizes noted, Hattie concludes that “it is less the content of the curricula that is important than the strategies (such as explicit teaching) that teachers use to implement the curriculum so that students’ progress upwards through the curricula” (Hattie, 2009, p.159).

An example of a specific investigation into the use of explicit teaching is found with Ryder et al. (2006), who employed a three-year longitudinal study examining the effects of explicit teaching pedagogy on: reading achievement, teacher perceptions, nature of classrooms, and referral rates for special education services. Structured interviews were followed by unstructured opportunity to respond so all teacher perceptions were noted and not just those elicited by the researcher questions. The authors also employed teacher questionnaires, and teacher observations in classrooms. Twenty-four classroom behaviours were listed including materials used within the class, discipline and management, teacher demeanour, and student response to instruction. Participants provided responses in the form of a Likert scale 1-5 and mean ratings were calculated across Autumn and Spring. Correlational analysis was conducted across the mean on the observational scale and school students’ results on a standardised test. When the characteristics of explicit teaching were evident and used by teachers, reading achievement improved and referrals to special education services decreased.

Kretlow and Helf (2013) surveyed 534 teachers about the teaching practices they used when teaching reading. Findings indicated that some teachers used commercial reading programs judiciously with elements of explicit teaching embedded in their practice; while others selected, adapted or ignored the requisite components of explicit teaching based on personal preference, perceptions of student needs or perceived learning style. These findings were based on teachers reporting their own practice.

Hattie (2009) reported results when teachers ignore explicit teaching characteristics and instead focus on other personal preferences such as ‘student learning style’. He reported effect sizes of instruction based on ‘aptitude–treatment’ interactions to be very low ($d = 0.19$), and even lower was ‘student control over their own learning’ effect size ($d = 0.04$).
Explicit teaching has been shown to have positive results when teaching students prosocial skills as well as academic skills. Cashwell, Skinner, and Smith (2001) illustrate how prosocial behaviours were taught to Year Two students. The peers of students with social skill difficulties were taught to report the prosocial behaviours they saw in their classmates. It was the use of interdependent reinforcement and progress feedback for the peers that guaranteed high levels of reporting prosocial behaviours were maintained. Thus the environment was altered to maintain results that had been achieved through explicit teaching of desired skills to the peers. Hattie (2009) concurred that there is a need to provide ongoing training on a regular and sustained basis when students are lacking these skills. With success in social skill competence, there were subsequent enhanced opportunities for learning and less disruption in class.

2.2.1.4 Issues and challenges associated with explicit teaching.

There are several challenges associated with the use of ET including: inconsistent definition of the practice, gaining the expertise required to implement the pedagogy with integrity, and maintaining implementation integrity to reduce the research to practice gap.

One of the difficulties in defining the strategy is that the terms *explicit teaching* and *direct instruction* are often used interchangeably in the literature (Pace, 2011). Hattie (2009) suggested that direct instruction has a bad name for the wrong reasons. The tenets of a very specific program (Direct Instruction, or DI) employ direct instruction methods which originated from the research of Becker, Engelmann, Carnine, and Maggs (cited in Ryder, Burton, & Silberg; 2006) in the 1960s and continues with strength through institute conferences today (NIFDI, 2016). The distinguishing features between DI and *direct instruction or explicit teaching* include: DI utilises students responding in unison, scripts are provided for the teacher to use, and the curriculum materials are tightly controlled. DI is described as the particular strategy that employs very ritualistic teacher practices and materials whereas the direct instruction or explicit teaching methods described by Rosenshine (2012) can be employed with any content and do not necessarily involve scripted teacher directions. The stylised DI program is confused with the more common approach to classroom instruction referred to as direct instruction (small’d’) or teacher directed instruction. Authors differentiated between explicit teaching and DI (Ryder et al., 2006) by
explaining that the approaches represented points along a continuum from most teacher-directed and task analytic to more student directed.

A further issue is the expertise required to actually implement ET with research-based characteristics identified by earlier meta-analyses and research studies. Bloom (1984) reported findings that suggest only 20% of school students thrive under conventional instruction; the remaining 80% do relatively poorly. This could be seen as a direct result of unequal treatment of students in classrooms and the lack of use of research-based characteristics of ET noted above. The higher achieving students receive the majority of teacher attention, feedback and reinforcement (Brophy & Good, 1970, cited in Bloom, 1984). Explicit teaching and mastery learning techniques require teachers to monitor practice and check understanding of all students. It requires more active teacher engagement with the class and checking for understanding than the traditional lecturing. In so doing, they are able to provide much more opportunity for students, including the 80% of students who may not succeed under conventional classroom instruction (Cazden, 1993).

Rosenshine (1986) described ET skills as a necessary part of the repertoire of an effective teacher and described steps that can be employed to teach the characteristics of explicit teaching to pre-service teachers. While the research indicates there is no one strategy that suits all students (Ryder et al., 2006), it is clear, however, that researchers over the past 30 years have identified features of teacher-led practice (described here) that positively affect student progress (across students groups, abilities, ages and curriculum areas). Once expertise with these features is achieved by pre-service and in-service teachers, ongoing monitoring and support is required to ensure the maintenance of implementation fidelity and ultimately a lessening of the research to practice gap, especially concerning students who have additional needs.

The second pedagogy of focus in this thesis is Cooperative Learning (CL). Following is a definition of the pedagogy, research support, issues and challenges associated with the use of CL, and implications for the current discussion.
2.2.2 Cooperative learning (CL).

2.2.2.1 Description of cooperative learning.

Cooperative Learning (CL) describes a set of instructional methods where students work in small mixed ability groups for the purpose of learning. Students are responsible for their own learning and for helping their team mates learn (Stevens & Slavin, 1995a). Wolford, Heward, and Alber (2001) define CL as a peer mediated instructional arrangement where small groups of students work together to achieve group success. Students are often assigned roles for completing a range of tasks (Magnesio & Davis, 2010; Malmgren, 1998; McMaster & Fuchs, 2002; Mercer & Mercer, 1998).

CL groups may be differentiated from traditional learning in that: traditional learning encourages individual students to compete for limited rewards within the classroom. Cooperative learning encourages interdependence with peers for mutual reward and works the following way, “If kids do not believe they sink or swim together the cooperation is not working” (Johnson & Johnson, 1992, p. 554). CL provides an alternative to the traditional competitive environment, increases instructional time, and enables teachers to further individualise instruction for students that need such adaptations (Kyndt et al., 2013).

Group work may be as simple as having students sit together to help each other, or may be complex organisational groups involving heterogeneous groups of purposefully selected students. At a minimum, cooperative learning group work needs to have the *sink or swim together* as a distinguishing feature to create the interdependence among participants (Magnesio & Davis, 2010). Many different CL arrangements have been developed and researched, such as those by Robert Slavin: Student Teams Achievement Divisions (STAD); Student Team Learning (Slavin, 1990); Team-Assisted Individualization (TAI); Cooperative Integrated Reading and Composition (CIRC); and Jigsaw II (Aronson, Stephan, Stikes, Blaney, & Snapp, 1978; Slavin, 1983), and in addition, those by the Johnson’s such as ‘circles of learning’ (Johnson & Johnson, 1992; Johnson, Johnson, & Holubec, 2009). Numerous other versions of cooperative learning have been developed over the years: Teams-games Tournaments (DeVries & Edwards, 1973); Group Investigation (Sharan & Sharan, 1992); Cooperative Structures and Numbered Heads Together (Kagan, 1989-1990); Learning Together (Johnson & Johnson, 1999); Cognitive Engagement in
Cooperative Learning (CECL) (Howard, 1996), and Complex Instruction (Cohen, 1994).

Even though there are various forms of CL, they all have common characteristics that are required for student success. Most authors agree on the following essential elements: task structure; group goals; individual accountability; interdependence; and group reward (Bain & Parkes, 2006; Fore, Riser, & Boon, 2006; Johnson et al., 2009; Malmgren, 1998; McMaster & Fuchs, 2002; Stevens & Slavin, 1995a). When the following elements are incorporated to introduce a CL task, effectiveness improves: clear lesson objectives; explanation of the task; group monitoring, and careful evaluation of student performance (Hattie, 2009). Studies by both Gillies and Ashman (2000, p. 259) and Wolford et al. (2001) described the use of effective teaching elements such as modelling, role play, corrective feedback and independent practice as essential parts of the CL process as well.

McMaster and Fuchs (2002) investigated the effectiveness of research on CL carried out between 1990 and 2000. They identified no less than seven different approaches or variations to CL, either in combination with other instructional methods or as a stand-alone. The aim of their review was to determine important characteristics of CL that have been signalled in the past including: CL and strategy instruction combined; CL and cross age tutoring; Cooperative homework teams; Learning together approach; CL and school wide restructuring; and structured versus unstructured tasks. Many studies were omitted from this review because of methodological issues such as: small sample size; short treatment durations; elements of the CL strategy not being reported thoroughly; and not specifying the components of CL. The results of the review supported claims that essential elements included group goals and group reward in line with individual accountability, and the careful structuring of the task activities. Across all the various forms CL may take, the most researched elements that promote optimal conditions for effectiveness included: positive interdependence created by group goals; individual accountability; and task structure adaptations (Hattie, 2012). These provide a powerful rationale for the use of CL for students who have difficulty learning.
2.2.2.2 Meta analyses that provide a rationale for the use of cooperative learning.

As a teaching pedagogy, CL is beneficial for students at all levels of ability as it enables teachers to increase instructional time and provides students with additional opportunities to respond and receive feedback (Antil et al., 1998; Beesley & Apthorp, 2010; Hattie, 2009; Kyndt et al., 2013).

Hundreds of studies have been carried out to determine the effect of CL on student achievement across every major subject, at all grade levels, and in all types of schools (Bain, 2007; Gillies, 2002; Hattie, 2009, 2012; Slavin et al., 1996). Kyndt et al. (2013) conducted a meta-analysis and found 54 articles yielding 121 findings. Results showed a positive effect of cooperative learning on achievement ($d = 0.54, p < 0.001$), attitudes ($d = 0.15, p < 0.01$) and perceptions ($d = 0.18$). Study domain, student age, and culture were moderators for the positive effects of cooperative learning.

Hattie’s (2009) meta-analyses of cooperative learning across 306 studies and 24,025 participants, found an effect size of $d = 0.41$ in favour of CL over other pedagogies. Various comparisons between the different CL characteristics resulted in the following findings in favour of CL: cooperative learning versus heterogeneous classes showed an effect of $d = 0.41$; cooperative versus individualistic learning $d = 0.59$; and CL versus competitive learning $d = 0.54$.

Slavin et al. (1996) reviewed 99 studies on CL. Of the 64 studies that used group rewards based on the sum of individual learning, 78% resulted in significantly improved gains; and none found negative effects. With an effect size of $d = 0.32$, these results showed that 32% of a standard deviation separated CL and control groups that did not employ group based rewards. These results were supported by the work of many other researchers (such as Haystead & Marzano, 2009; Johnson & Johnson, 1989; Marzano, 1998), suggesting that an interpersonal reward structure encouraged group members to support each other in learning. If group goals in the task can achieve a moderate effect size and group rewards can achieve a similar result, perhaps the combined effect of these two elements could prove quite powerful (Bloom, 1984; Hattie, 2012). Marzano et al. (2001) reported an overall effect size of 0.73 when group goals and group rewards were incorporated into the design. A later meta-analysis by Beesley and Apthorp (2010) which involved over 2,000 students across multiple
grades, subject areas, and various measures of academic achievement, found a composite effect size of 0.44 indicating an average gain of approximately 17 percentile points.

Other authors have compared the effects of CL from within the same study (for example, Fantuzzo et al. (1992, as cited in Slavin, 1996). Fantuzzo and colleagues made a direct comparison between rewards alone, and the larger CL structure alone, and found the combination of reward use and structure differentiation to be by far the most effective. The effect of reward plus structure conditions on achievement resulted in an effect size of $d = 1.42$; and reward alone had a larger effect than structure alone. As with the suggestions proposed by Bloom (1984) regarding explicit teaching variables, the effective variables used in combination may in fact have a cumulative and much higher impact than the variables in isolation. This notion highlights the important variables that need to be included in any variation of CL that is used.

2.2.2.3 Specific research support for cooperative learning.

Examples of CL as a focus in the research literature indicated it has been used extensively in education, for example, in the area of curriculum achievement: reading (Fore et al., 2006); development of vocabulary skills (McMahan, 1993); mathematics (Xin, 1996); and also in cognitive domains such as generic higher order thinking and self-esteem (Jenkins, Antil, Wayne, & Vadasy, 2003). The research also pertains to supporting inclusion of students with special needs in regular classrooms (Bryant & Bryant, 1998; McMaster & Fuchs, 2002; O'Connor & Jenkins, 1993; Putnam, Markovchick, Johnson, & Johnson, 1996; Wolford et al., 2001).

Mainzer (1993) investigated the use of CL by comparing students with mild intellectual disability in separate classes to those who were mainstreamed. One group of students received instruction using CL pedagogies and the other group received more traditional methods of instruction. The results showed the mainstreamed students exposed to CL had significantly higher gains in reading vocabulary and comprehension as well as language expression than those students with mild disabilities taught using traditional methods.

When investigating the effects of interventions for students with learning disabilities, Swanson and Hoskyn (1998) provided a summary of effects from group and single-
subject designs. They found a mean effect size of $d = 0.56$ for group designs and as high as $d = 0.9$ for the single-subject designed studies. The researchers concluded that a combination of explicit teaching characteristics, along with cooperative learning pedagogies, were the most effective for students with learning disabilities. They stressed the importance of using both approaches (ET and CL) to maximise the effect on student achievement and that CL groups provided the best opportunity for multiple and deliberate practice of skills being learned.

In spite of the apparent plethora of research available supporting the use of CL, Gillies and Ashman (2000) suggested there is still insufficient empirical work involving students with learning difficulties. This is an interesting observation given the inherent ability CL offers teachers to deliver large scale individualised instruction in the group setting of a classroom (Beesley & Apthorp, 2010; Blank, 2013; Slavin & Lake, 2009; Tomcho & Foels, 2012).

What begins to emerge from the literature is that properties of CL are not only being widely investigated in terms of *curriculum achievement*; but also in terms of the actual *processes involved in group work* as a means of inclusion. Wolford and her colleagues (2001) were interested in the group interaction skills of students with learning difficulties (LD) in a study involving four high school students with learning disabilities. They chose independent variables of: rate of recruiting peer assistance, rate at which students received praise and feedback, and productivity and accuracy of work completion. Their focus was on interaction skills needed for CL as well as task completion. Appropriate recruiting was classified into “appropriate” and “inappropriate” variations; peer assistance and attention was classified in a similar way. Academic productivity was classified according to the number of items completed correctly on standardised class assignments set. A multiple baseline across subject design was employed. After training, all four participants recruited peer assistance the required amount of times at levels of 2 to 3.4 times per minute across the generalisation and maintenance stages of the study. The number of inappropriate recruiting responses decreased. Peer assistance administered increased at a rate of 1.5 times higher than during the baseline phase. The fact that assistance continued to be sought after all teacher prompting and consequences had been withdrawn was a positive finding.
The authors reported that a functional relationship existed between training students to recruit assistance and increased academic productivity. They ranged from 50% completion during baseline to 80% completion after intervention. This study reinforced the notion that peers can be an important source of feedback and positive attention during CL in the inclusive classroom. Teaching students to recruit instructional feedback increased many times the potential of communities of reinforcement in a typical classroom, and affirmed that students with LD can engage in the behaviours required for successful CL activities (Wolford et al., 2001).

Wolford et al. (2001) offered the relevant warning that it would be a mistake to assume students with learning disabilities will function successfully in CL groups without sufficient training and support in group processes. The objective is not just to master curriculum content, but as noted above by Woolford and her colleagues (2001), the CL group processing skills also need to be explicitly taught. It is imperative for interaction skills to be explicitly taught and not to assume that simply placing students in groups, even with a well-structured task, is sufficient to guarantee success.

Gillies and Ashman (2000) also investigated group processing skills and how students with disabilities interacted within CL groups and the impact of this interaction on their behaviour and learning outcomes. They investigated the effects of CL on 152 Grade Three students over three school terms while working in the curriculum area of social studies. Students were randomly assigned to either structured or unstructured groups and measures were taken of student learning, interactions, and behaviours. The structured group condition involved two training sessions to develop small group and interpersonal skills (as identified by Johnson et al., 2009). Extensive descriptions of the training involved in small group interpersonal behaviours were provided, including what the behaviours would have looked and sounded like. The unstructured group was allowed the same amount of time to discuss group behaviours but no specific training was provided. Group observations included: cooperative behaviour, non-cooperative behaviour, individual task-orientated behaviour, and individual non-task behaviour. These were all presented in observable terms and reliability of observations was as high as 85-95% across experienced observers using video recordings of sessions.

Findings indicated an impact on ‘non-task’ behaviour by the time the second data collection sample was taken where the non-task behaviour had significantly decreased
for students who had LD more than for those in the structured task condition. Those in the structured condition also obtained higher post-test results with the comprehension questionnaire showing an effect size of $d = +1.43$. Using standardised tests, the learning outcomes were considered to be *educationally* significant and not *statistically* significant; the authors suggested the lower result was due to the use of a standardised test to judge results and said that the test may not in fact have been sensitive to the curriculum being used. This is often a difficulty associated with the use of standardised testing (Taylor, 2000).

In 2002, Gillies repeated the study to determine if there were any residual effects of CL training two years after the initial study. Gillies (2002) measured student behaviour states associated with cooperation such as task orientated behaviour, active attention to others, and verbal interactions of the comparison groups. Findings confirmed that after two years, students who had received training were able to demonstrate cooperative behaviours and verbal interactions significantly more than those who had not received the training.

### 2.2.2.4 Issues and challenges associated with cooperative learning.

Issues based on the CL research conducted to date include: methodological issues and study design; target audiences for CL; effect sizes based on norm referenced as opposed to criterion referenced tests; and fidelity of implementation, and scale up of CL use following successful studies.

Even though many studies report successful results, McMaster and Fuchs (2002) warned that mixed results were more likely to be the norm due to methodological issues in study design. They reported on results by Tatyema (1990, cited in McMaster & Fuchs 2002) that showed only 50% of studies could claim statistically significant results and posited that this figure was highly dependent on the *type* of CL being used. McMaster and Fuchs (2002) followed up on Tatyema’s work and reviewed studies from 1990-2000 with strict inclusion criteria. Of the 15 studies that were narrowed down for inclusion in their review, only *seven* reported statistically significant results (Brandt & Ellsworth, 1996; Gillies, 2002; Gillies & Ashman, 2000; Jenkins et al., 1994; O'Melia & Rosenberg, 1994; Stevens & Slavin, 1995b). They also cautioned that individually administered criterion referenced tests are more likely to produce large
effect sizes than group administered norm-referenced tests. This was confirmed by Gillies and Ashman (2000) and Wolford et al. (2001).

Some authors suggested that high achievers may be held back when involved in CL, however, having to give elaborated explanations in cooperative learning helps those who give them, more than those who receive them (Malmgren, 1998). Most studies included in the Slavin et al. (1996) review found equal benefits for high, medium and low achievers in comparison to control group participants.

In terms of ‘getting to scale’ with the use of CL in classrooms, many studies have documented the success of reading in the nationwide program known as Success for All (SFA), which used cooperative learning in disadvantaged elementary schools (Slavin et al., 1996). Results indicated the gains for the SFA schools were greater than for the state, and SFA schools were increasing further in advantage with each year of implementation. Jones, Gottfredson, and Gottfredson (1997) evaluated SFA in comparison schools across Carolina and their results yielded mixed findings with some significant gains and some negative. A number of North Carolina schools voted not to continue with the program and it was standard to only bring the SFA program to schools where 80% of staff were on side. Poor interpersonal relations within a school may have acted to sabotage potential gains to be made (Fore et al., 2006).

When considering maintenance and the scaling up of the CL strategy, Slavin et al. (1996) reported that 70% of primary school teachers made some sustained use of CL. However, as noted in earlier discussion, this 70% may actually be a misrepresentation and may not include high fidelity implementation given the wide interpretation of what CL actually entails.

Using a case study of 12 practicing teachers, ranging in experience from one to 26 years, Hennessey and Dionigi (2013) set out to determine the teachers’ knowledge of cooperative learning terms and function within their classrooms. Following analysis of data collected from semi-structured interviews with the participants, they noted that half the participants were categorised as having a limited understanding of cooperative learning principles and pattern language terms, four had a general understanding, and only two had a detailed understanding of the pedagogy. The results suggested there is a
need for teacher training and continued support for implementation of research-based pedagogies in schools.

In 1998, Antil et al. examined the prevalence, conceptualisation, and form of cooperative learning used by 85 teachers from six elementary schools. The participants responded to a survey, and interviews were conducted with a subset of those teachers \( (n = 21) \), all of whom indicated having daily cooperative lessons across several curriculum areas. The majority of teachers subscribed to cooperative learning to achieve both academic and social learning goals, structured tasks for positive interdependence, and taught students skills for working in small groups. The study authors applied criteria for cooperative learning derived from the research literature and found that few teachers were employing recognised forms of the cooperative learning practice, that is, they did not tie individual accountability to group goals.

In 2003, these authors reanalysed the same data set from the 1998 study (Jenkins et al., 2003), this time with a focus on students with additional needs. Teachers were generally positive about the use of CL with these students. The major benefits they identified were: improved self-esteem, providing a safe learning environment, and giving a greater voice to the students. The authors noted that these findings did not match the research focus on student achievement. In fact, better learning was ranked fifth by the participants. The main classroom modification made for students with additional needs was ‘selecting suitable partners’. In general, teachers expected the students’ teammates to make adjustments to the task to assist learning for the student with additional needs. The authors concluded that the omission of individual accountability may be the result of insufficient training in its importance. If CL and all of the necessary characteristics are not implemented with fidelity, the efficacy of the approach would be compromised (Jenkins et al., 2003).

Slavin et al. (1996) reiterated the need for effective professional development on CL to support high quality and sustained implementation. He suggested comparisons between: school-wide and teacher implementations; expert versus peer coaches; generic principles versus specific strategies; and the use of teacher learning communities. As noted earlier by Gillies and Ashman (2000), there were inherent difficulties when using standardised assessments to measure students’ progress. The
assessment of CL should incorporate a curriculum-based measurement (CBM) that is more sensitive to content in the curriculum.

In summary, this section has described the two key pedagogies that will be the focus of the intervention in this study. It is important to acknowledge that ET and CL are not the only two pedagogies with research efficacy available to teachers. However, there was compelling rationale for their use in a study of pre-service teachers’ inclusive practice given they are widely recognised and supported as core practices for the inclusive classroom. The results from meta-analyses, such as Hattie (2009, 2012) and Marzano (1998), have shown that ET and CL were the approaches that had extensive research support in schools when particular characteristics were included. The next section will focus on implementation fidelity and sustainability of these approaches.

2.2.3 Measuring implementation fidelity, sustainability and scalability of explicit teaching and cooperative learning.

While there is a strong base of evidence in applied experimental research for the use of ET and CL, a need exists to examine these approaches in routine practice. Success has been demonstrated in controlled experimental research over time, settings and populations, (e.g., Fraser et al., 1987; Gillies & Ashman, 2000); what was still needed from research is demonstrated sustainability and scalability of these practices. Studies that investigated implementation fidelity for classroom settings were located and these are elaborated upon in the following discussion.

Klingner, Vaughn, Arguelles, Hughes, & Ahwee Leftwich (2004) examined factors that led teachers to be high, medium or low implementers of research-based pedagogies such as explicit teaching to teach the skills of reading. The authors evaluated implementation integrity three years after the initial training occurred and thus tested the sustained use of these pedagogies. Findings delineated factors that affected the high implementers, for example: they considered the impact of the pedagogies on their students, were able to persist with new methods, could problem solve, and could also determine the overlap between the existing and new school initiatives. In 2004, Klingner et al. delineated other factors that positively influenced sustainability. For example: “Teachers need to see concrete examples of how a theory relates to their students and their circumstances” (p. 413). This is not surprising given the nature of ET which uses modelling, guided practice and independent practice as
elements of the pedagogies when teaching students. Why would the same not apply to adult learners?

When teachers were faced with too many changes and discord between new and existing knowledge about effective pedagogies, it became easier to revert to familiar patterns of behaviour, even if these were less effective. Findings from the same study indicated that the medium level implementers were frustrated by the range of ability levels in the class (Klingner, et al., 2004). Clearly there was a need for these teachers to be taught effective pedagogies (that is: the means of differentiating the strategies that had been the focus of their in-service training) to a point of mastery, maintenance and certainly generalisation.

Other studies that examined the routine use of these approaches in classrooms showed implementation to be highly variable (Cook & Tankersley, 2012; Kretlow & Helf, 2013). These authors found teacher understanding of evidence-based pedagogies to be inconsistent, as was the implementation from class to class in schools. Few practicing teachers appeared to employ those recognised forms of the practice that are associated with positive achievement effects (Hennessey & Dionigi, 2013). This is a disconcerting finding given that the assumptions made about the benefits of these practices are based upon research studies where their procedural features were implemented with high degrees of integrity (Slavin & Lake, 2009). This research to practice gap called into question whether these practices were being used in ways that were actually responsive to individual difference in classroom settings (Grima-Farrell, 2012; Grima-Farrell, Bain, & McDonagh, 2011). This finding also pointed to the need to prepare teachers to deploy the research-based characteristics of inclusive practices with integrity in their classrooms.

Little and Houston (2003) investigated processes to reduce the research to practice gap in schools. These particular authors provided a conceptual framework as well as specific implementation activities to ensure best practices were used and implementation was assessed empirically. The process proposed included the following four steps:

a) Identify research-based instructional pedagogies such as ET and CL;

b) Select teams of teachers to attend the professional development about these pedagogies;
c) Implementation of the instructional practice in the classroom from initial training to checking fidelity of implementation; and

d) Data collection of the results of students learning through traditional and action research methodologies.

Each of these steps specified observable criteria that could be used to judge reliability and validity. For example, for selection of research-based articles about practice (in step one), reliability scores of .85 - .92 were achieved against the selection criteria. Step two involved teams of teachers: A team selection criterion was developed to ensure that only committed personnel would be involved and so that participants knew what they were going to be involved in. Step three involved classroom implementation: Training involved modelling and guided feedback as well as providing necessary materials. Electronic and mail networks were established for this purpose and treatment validity was established. Critical teaching behaviour checklists were also developed for the different pedagogies being implemented. Step four involved data collection which was completed on student impact data and progress as well as on the implementation process. Multiple measures were considered for both (especially curriculum-based measures for student data). Action research was used as the vehicle to frame the process.

To pursue sustainability, participants were then able to volunteer to become ‘trainers’ and a criteria sheet was developed for considerations to be taken into account. One year after initial training, 20,000 students had demonstrably benefited from instruction. Mean growth for student achievement across academic and social domains involved substantial effect sizes from .67 - .89.

Unlike the route suggested by Denton, Vaughn, and Fletcher (2003) and Klingner et al. (2004), Little and Houston (2003) suggested starting with a state-wide approach as a means of achieving scale and sustainability. Deshler (2003) made policy recommendations to be considered by funding sources to assist bridging the gap between research and practice. He especially recommended an atmosphere of research that needed to be evident across all levels of education in order to make a sustained difference (Loreman, 2007).
Kretlow and Helf (2013) referred to a five–component framework devised by O’Donnell in 2008 for defining implementation fidelity. These steps were comparable to those of Little and Houston (2003). O’Donnell (2008) suggested that very few studies existed that examined any of these five components, let alone the combination of them.

Lopata, Miller and Miller (2003) reviewed actual versus preferred use of CL characteristics among exemplar teachers. The actual use of the desired characteristics of CL fell significantly short of the levels teachers would prefer to be using. Even though teachers were aware of the implementation fidelity they should be using, various reasons prevented it from occurring. In the study, the use of coaching was incorporated and the efficacy of its use tested using a multiple baseline design. Results indicated a functional relationship between coaching and improvement implementation. This finding echoed step three offered by Little and Houston (2003): “classroom implementation of instructional practice following initial training to check maintenance of implementation fidelity” (p. 78). This improvement was sustained over a maintenance phase of the study and was also reflected in student results.

As noted earlier, a more focused set of inclusive education indicators was called for as opposed to the broad indexes supplied by Hueng (2006). Cook and Tankersley (2012) suggested first identifying the practices to use and then providing a guiding framework for implementation. The structure of steps elaborated by Little and Houston (2003) provided such a framework, including support from initial professional development right through to checking implementation fidelity using critical teaching behaviour checklists. Student impact data was collected as the final step and the whole process was continued through use of an Action Research paradigm. Lee and Picanco (2013) suggested a way forward to avoid teachers having ‘pedagogical fixations’ due to philosophical orientations by tying research-based pedagogies to the phases of learning within which the students are located. For example, ET would be required for the acquisition phase of learning, while CL would be more appropriate in the generalisation phase. This could be supplemented with the differentiation proposed by Tomlinson (2001) through the areas of content, product and process (Ernest, Heckaman, Thompson, Hull, & Carter, 2011).
The combination of these processes would have the potential to assess current practices within classrooms as well as to develop teacher skills. No similar research was located about implementation fidelity in PST programs. The considerations described in studies of sustainability and research to practice were employed to inform the design of this current research, in particular, in the way in which implementation integrity is measured. The case for the use of explicit teaching and cooperative learning is strong as the pedagogies have extensive research support and are consistently identified as important for inclusive practice.

As well as the ability to deploy evidence-based pedagogies, sustain their use, and implement them to scale with integrity, collaboration skills utilising consistent professional language also needed to be investigated as potential factors that mediate successful inclusion. Consistent professional language used for collaboration will be discussed in the following section.

![Figure 2.3](image)

**Figure 2.3.** The capacity to communicate and collaborate about student learning using a professional pattern language.

2.3 *Teachers’ Capacity to Communicate about Learning using Professional Pattern Language*

One explanation for the lack of implementation integrity in classrooms could be the lack of consistently used professional language that described essential characteristics of the pedagogies being employed. For example, teachers may use terms with differing meanings and have differing understandings when discussing pedagogies used in their classrooms. This lexicon was frequently presumed to exist in communities of practice in educational settings, although the findings derived from the longitudinal study of
schools (Bain, 2007; McLaughlin & Talbert, 2001) indicated that the presence of such a consistently used lexicon was much more likely to be the exception than the rule.

Use of a common professional or pattern language in collaboration with colleagues is important to inclusive educators as a means to solve problems and address student learning needs at a sophisticated level. A shared pattern language aids the ability to provide feedback about teaching approaches, student need, and decision making. If a community is to evolve in the practice of collaboration, it first requires its own pattern language which consists of the terms the community uses to express its schema (or organisational framework) and to organise its models and practice (Smethurst, 1997).

For a teaching community of practice to be the social container for genuine professional interaction, all teachers require the knowledge associated with the teaching and learning approaches valued by the community. For example, if inclusive educators are to work together to solve a problem related to the use of cooperative learning or explicit teaching, they all need to understand the terminology described in the literature, as well as the roles and goals of those practices within the broader context of the community’s overall schema of inclusive practice. The schema requires a pattern language used to express the essential features of those teaching and learning approaches. Pattern language is required as a basis, then as a means, to collaborate with others at a sophisticated level. It becomes the term of reference and method for metacognitive understanding whilst collaborating (Bain, 2007).

The findings of Hennessey and Dionigi (2013) are educative here. Their work found that overall, teachers have low understanding of those terms that would constitute professional pattern language related to cooperative learning. These authors surmise that “Teachers require a particular pattern language in order to use cooperative learning effectively” (p. 52). The data were based on self-report rather than actual observations of implementation, however the authors concur about the need for teacher training and continued support for implementation in schools. The use of professional pattern language would promote the development of conditions for more effective collaboration. Unfortunately, as Darling-Hammond and Richardson (2009) noted, schools do not provide the ongoing assistance necessary to implement and maintain such approaches.
Similar findings were reported in the studies by Antil et al. (1998) and Jenkins et al. (2003), where the researchers noted that 76% of teachers in the sample used a ‘soft’ form of CL; one that did not include procedures recommended by researcher-developers to establish individual accountability, which entails computing individual scores and rewarding group performance. Many teachers embraced the use of the teaching pedagogy for their students, however, the omission of ‘individual accountability’ and group goals made it less efficacious in terms of student achievement. The requisite terms used to describe CL were in the teachers’ vocabulary sets however were not used in a way that enhanced implementation fidelity of CL.

A study by Bain, Lancaster, and Zundans (2009) set out to determine whether PSTs in an inclusive education teacher preparation course were able to deploy pattern language related to specific inclusive teaching pedagogies. The study sought to establish whether the frequency and sophistication of pattern language use increased as the pre-service course progressed and as PSTs learned more about inclusive approaches. In the Bain et al. (2009) study, frequency of pattern language was measured by counting the terms used after participants wrote a reflective piece about their pedagogies. The sophistication of pattern language was measured by applying the SOLO taxonomy developed by Biggs and Collis (1982). The results indicated that pattern language frequency and sophistication co-varied with participation in the course, and increased over time. The study showed that it was possible to successfully teach pattern language in a PST setting.

The requirement for a consistent pattern language would seem to covary with the need for collaboration between staff. Note that collaboration is not an independent variable in this study but as a platform for using sophisticated language when solving complex problems that arise with students who have additional needs it warrants brief discussion here. Collaboration is defined as an interactive process that enables people to generate creative solutions to mutually defined problems that need to be addressed (Idol, Paolucci-Whitcomb, & Nevin, 1994). Successful collaboration includes specific facets: a) parity of group members, b) mutual goals, c) shared decision making and responsibility, d) shared resources and accountability, and e) valuing the opinions and expertise of others (Friend & Cook, 2013). Widely used steps include: delineating problems, brainstorming solutions in a divergent manner, prioritising solutions in a
convergent manner, agreeing on a final goal, making a plan, and evaluating in a timely fashion (Friend & Cook, 2013).

Collaboration skills involve the ability to engage in effective communities of practice to support problem solving in complex learning environments. Over the last twenty years, collaboration techniques have become a cornerstone of inclusive education practice used to develop and review individual education plans, for instructional problem-solving, as a medium of engagement with parents, and by the different professionals who work with students with inclusive needs (Friend & Cook, 2013; Salend, 2005). Collaboration is frequently identified as a key to the successful conduct of inclusive classrooms and schools with this positive relationship well documented in the literature (Dettmer, Knackendoffel, & Thurston, 2013; Hamman, Lechtenberger, Griffin-Shirley, & Zhou, 2013; Loreman, Deppeler, & Harvey, 2005). Research suggests that collaboration and collaborative problem-solving should be on the agenda of all schools and teacher education courses.

The role and processes of collaboration have also been connected to the related construct of communities of practice (Wenger, 2000). Communities of practice (CoP) have been advocated in inclusive education to tap expertise and bring stakeholders together to problem-solve and communicate professional knowledge (Buysse, Sparkman, & Wesley, 2003; Linehan, Muller, & Cashman, 2005). They can be viewed as the places where the instrumental process of collaboration and collaborative problem-solving are embedded systemically in a local context. Like collaboration, the construct of CoP also resonates powerfully with the challenges of inclusion (Wesley & Buysse, 2001), and especially the need for school-wide teacher collaboration if the inclusion of students with additional learning needs is to be successful. The processes of collaboration require: the elimination of traditional role boundaries among staff in order to achieve equity; reciprocity and mutual respect (Friend & Cook, 2013); and the establishment of collegial norms for interactions (Salend, 2005).

Van Garderen, Stormont, and Goel (2012) provided an extensive review of collaboration between general and special educators. They noted that collaboration had been recommended for several years as a means of improving student performance; however, very little empirical evidence existed about its success.
Several researchers have conducted meta-analyses of collaboration (Murawski & Swanson, 2001; Reinhiller, 1996; Scruggs, Mastropieri, & McDuffie, 2007; Welch, Brownell, & Sheridan, 1999). Murawski and Swanson (2001) and Scruggs et al. (2007) conducted meta-analyses on co-teaching and found that the practice had an overall effect of 0.4 for influencing student outcomes. They cautioned that the findings should be considered tentative as there were only six articles in their analysis and the literature search was limited to the years 1989-1999. Team teaching and problem solving was investigated by Welch et al. in 1999. Only Murawski & Swanson (2001) concentrated on student outcomes as a result of teacher collaboration.

In a study by Klingner et al. (1999), the authors noted that the participants who implemented strategies with integrity persisted as long as they could problem solve with others and be provided with feedback about their practice. These findings were echoed by Bain (2007) when investigating a process of school reform to determine evidence of faculty collaboration and culture. Bain sought to compare the views on collaboration and school culture of a school faculty that applied a model of self-organisation (SOS) with those of the faculty in 42 other schools. This study investigated collaboration in a school environment where collaborative process and inclusive practice were applied across the entire school. The data collection used the Research for Management (RSM) interview form (Buckalew, 1994 cited in Bain, 2007) to collect data across three phases of the SOS project, and compared results to findings from Buckalew’s own five-year international study. Results showed median scores in each phase of the SOS project to be higher than those from the RSM study. Teachers were unanimous in attributing the higher scores to the collaborative problem solving model used in the SOS design.

Thomson and her colleagues (2003) discussed a consortium of three universities in New Zealand who formed with the intention of developing a program to ensure the effective acquisition of collaborative consultation skills for their teachers working in schools. The impetus for the program development was to assist support teachers to transition from a ‘pull-out’ role to one of inclusion and collaboration with the classroom teacher. They effectively had to become agents of change using the collaborative consultation framework to implement their changed roles. Both support teachers and classroom teachers found the professional development beneficial in supporting student learning and behaviour.
More recently, Thomson (2011) conducted a qualitative study that investigated the service delivery model of collaborative problem solving in ways that assisted teachers, support teachers and students working in inclusive settings. She interviewed 14 teachers about their experiences with the collaboration process that were specific to the contexts and participants themselves. While collaborative approaches were well recognised as key features of inclusive practice, they needed to be advanced within the structure and conduct of schools if they are to be deployed effectively for the benefit of students.

The literature reviewed indicated that collaboration is required in order for teachers to problem solve and find solutions to difficulties as they arise in their classrooms (Buysse et al., 2003; Cochran-Smith, 2005; Linehan et al., 2005). Common shared professional language (pattern language) is required to engage with colleagues collaboratively in order to work in inclusive learning environments. Frequency and sophistication of pattern language could be measured in the same way that Bain et al. (2009) employed in their study. Just as reliability and validity of observational assessments could be checked for use of teaching strategies (Hintze, 2005), these principles could also be employed for observational assessment of the collaboration and pattern language processes. These considerations need to be taken into account for the current study to allow teachers to discuss inclusive pedagogies such as explicit teaching or cooperative learning with a level of understanding and sophistication.

When these skills of collaboration using consistent professional pattern language and pedagogical expertise are evident, it is expected that self-efficacy and competence of teachers working with students who have varying needs will be increased. Teacher efficacy is related to teacher behaviours in the classroom and is said to effect teacher levels of effort, planning, organisation, persistence and resilience. As well as sophisticated use of pattern language to discuss characteristics of pedagogies that make a difference to inclusion, attitudes, beliefs and self-efficacy of teachers also need to be improved as they work with students who have different needs (Avery & Meyer, 2012).
Figure 2.4. Positive self-efficacy about inclusion in schools.

2.4 Teachers’ Self-Efficacy in the Inclusive Classroom

The focus in this section is on teachers’ personal self-efficacy for research-based pedagogy use. It is thus task specific (focused on the use of explicit teaching and cooperative learning pedagogies) and context specific as it involves inclusive classroom settings.

A teacher’s sense of self-efficacy is defined as “the teacher’s belief in their capability to organise and carry out actions required to successfully implement specific teaching in a particular context” (Jamil, 2012, p. 6). Self-efficacy has its antecedents in social cognitive theory and involves efficacy beliefs directed towards a teaching context. These beliefs powerfully predict the choice of task that teachers take on, the effort expended, persistence shown, and ultimately the level of student success achieved (Bandura, 1989, 1997). Those with a strong sense of efficacy are thought to display the use of teaching characteristics that enhance learning. In a longitudinal analysis, Holzberger et al. (2013) found that the teachers’ self-efficacy in their study was related to instructional quality, and that cross-sectional correlations between self-efficacy beliefs and characteristics of instruction were substantiated.

Specifically, teachers with a high sense of self-efficacy utilise more behaviours that have the potential to enhance student learning and motivation (Abbitt, 2011; Hastings, 2012). Studies on implementation fidelity discussed earlier by Klingner et al. (2004) clearly state that the high implementers of research-based pedagogies were those who were concerned about student success. When their students succeeded, teachers’ levels
of self-efficacy were in turn improved. This led them to persist with new pedagogies and problem solving (Holzberger et al., 2013).

There has been a shift recently in the focus of self-efficacy studies from whether teachers can ‘cope’ in an inclusive classroom (Chiner & Cardona, 2013), to teacher self-efficacy about strategies to ensure success for their students (Ernest et al., 2011; Hastings, 2012; Holzberger et al., 2013; Viel-Ruma, Houchins, Jolivette, & Benson, 2010; Yilmaz, 2011). For example, Ernest et al. (2011) described a case study of a beginning teacher's use of differentiated instruction using Tomlinson's (2001) categories of content, product, process, and learning environments. After implementing differentiated instruction, the qualitative data indicated an increase of the PSTs’ self-efficacy when they realised that it was possible to meet the needs of all children in an inclusive classroom. Hastings (2012) discussed Bandura’s (1989, 1997) perceptions of efficacy based on four sources: enactive attainment; vicarious experience; physiological and emotional states; and verbal persuasion. Based on interviews and literature analysis, the findings indicated that enactive attainment and vicarious experiences were two sources of self-efficacy that need to be included in a focused manner in PST curriculum (Hastings, 2012).

After interviewing 54 teachers, Yilmaz (2011) determined that the more proficient teachers perceived themselves to be with pedagogical content knowledge, the more efficacious they felt implementing instructional pedagogies and classroom management. The pedagogical content knowledge that was the focus in this study was teaching English as a second language. One could supplant knowledge of inclusive pedagogies such as explicit teaching or cooperative learning and hypothesise about similar findings. As the authors suggested, studies conducted on the basis of teachers’ beliefs were important in determining the way teachers perceived and organised instruction and assessment in their classrooms. A teacher’s sense of efficacy directly influenced the kind of environment they created for their students to bring about learning (Yilmaz, 2011). That environment could include the pedagogies employed, the amount and depth of problem-solving using sophisticated language, and the extent to which there exists a genuine community of practice.

Using longitudinal quantitative methods Holzberger et al. (2013) designed a study which combined a self-report measure of teacher self-efficacy with teacher and student
ratings of instructional quality and 2-level cross-lagged structural equation analyses. Data were collected from 155 German secondary mathematics teachers and 3,483 Grade 9 students at two measurement points. The researchers found that cross-sectional correlations between self-efficacy beliefs and characteristics of instruction were significant. They were able to determine that a teacher’s self-efficacy was not only a cause of educational processes being used but also a consequence. In other words, when teachers successfully used pedagogies that incorporate research-based characteristics with their students that resulted in improved outcomes, their own self-efficacy also improved. It appears to follow that when research-based pedagogies are used in classrooms to enhance student learning, teacher self-efficacy was also enhanced.

In spite of the consensus about the importance of teacher self-efficacy, there were limitations to interpretation of the research presented. For example, there was no consensus about a clear definition or measures of the construct (Chan, 2008). The concept of ‘locus of control’ and Bandura’s social cognitive theory have been major sources of the initial conceptualisation of the construct. Hoy and Spero (2005) identified general collective self-efficacy, which explains that teachers’ beliefs as a profession do affect student outcomes. The other area identified was that of personal or teacher self-efficacy, which refers to teachers’ beliefs about their own abilities to effect student outcomes. Teachers do not feel efficacious in all teaching situations and their personal efficacy beliefs can be task and context specific. Few studies have sought to look at both perceptions of efficacy as well as actual practice.

Thus, researchers continue to grapple with the notion of different efficacy scales being constructed to measure different tasks in the complex job of teaching. Bandura’s scale included 30 items that covered seven categories of teacher behaviours. Tschannen-Moran, Hoy, and Hoy (1998) summarised these categories into three broader areas: instructional strategies, classroom management, and student engagement. Chan (2008) developed a scale to assess six domains of teacher self-efficacy: teaching highly able learners; classroom management; guidance and counselling; student engagement; accommodating diversity; and enriching learning. The scales certainly contain overlaps as well as discrepancies. Not all were focused on inclusive classrooms and, indeed, the scale development to measure the construct by Chan (2008) had only one dimension specifically focused on managing diverse learning needs in the classroom.
A need exists to investigate the mediating variables involved in enhancing self-efficacy to ensure that student outcomes continue to improve with a more dedicated focus on inclusive education and also to consider whether self-efficacy is situation specific within the context of inclusive practice. A scale that specifically encompasses the characteristics of research-based explicit teaching and cooperative learning would be beneficial for classroom use.

*Figure 2.5.* Translation of skills and attitudes for successful inclusion in schools.

It is clear from the literature that students depend on highly qualified teachers, who believe themselves to be efficacious, carry out efficacious interventions, and build system capacity to deal with change and solve problems that arise in complex learning environments. The question arises: Where are these people and how do they become highly qualified (King-Sears et al., 2004)? The research lends strong support to the processes and pedagogies described so far in the review. The section that follows will extend the discussion to pre-service teacher education programs which continue to be the vehicle for providing teachers with preparation for the inclusive classroom.

The subsequent sections of the literature review focus on:

- The ways inclusive education indicators are developed and evaluated in pre-service teacher education (PTE) courses; and finally
- The place for theory in course design.
2.5 Inclusive Education in Pre-service Teacher Education Programs

There is widespread support for teacher preparation in the area of inclusive education; however, there is also international concern about whether the preparation pre-service teachers receive for inclusion is adequate (Edelen-Smith, Prater, & Sileo, 1993; Snyder, 2012; Wiseman, 2012). Husebo (2012, p.445) describes one rationale for this concern where teacher education is often subjected to “intersecting demands and unattainable expectations” by accrediting bodies, and this is exacerbated by the ‘tokenistic’ inclusion of minimal courses of study. There continues to be a great deal of conjecture about whether pre-service teacher education does what schools require to ensure successful inclusive education (Snyder, 2012), such as ensuring that PSTs are able to use pedagogies that incorporate research-based characteristics and to work collaboratively with their peers using the professional language described in previous sections of this review.

Within the field of PST education more broadly, many authors agree there is a need to close the gap between theory and the ability of PSTs to implement research-based practices (Bauer et al., 2004; Brownell, Ross, Colon, & McCallum, 2005; Husebo, 2012). This would appear to be even more crucial in the area of inclusive education. For many Australian PSTs, their only exposure to the field of inclusive education is participation in a mandatory introductory course included in their undergraduate teacher education programs (Carroll et al., 2003; McRae, 1996; Sharma, Forlin, & Loreman, 2008). Florian and Rouse (2009, p. 711) suggest that separate courses for inclusion are often marginalised within general education programs or presented as discrete content. Accordingly, teacher education programs have been accused of having ‘myopic vision’. Gill and Chalmers (2007) implied that the use of single units on diversity within an entire PST program means that inclusion and diversity are addressed in a very tokenistic way and they call for validation of the current practices as opposed to tokenism.

A further indicator that pre-service teacher education is problematic is the high attrition rate amongst new graduates, and the even higher attrition rates amongst inclusive special educators (Brownell et al., 2005), up to 2.5 times higher than other new educators (Billingsley, 2010). One explanation for the high levels of graduate teacher attrition is said to be the perceived mismatch between pre-service preparation and the actual working conditions of teachers (Dempsey et al., 2002; Whittaker, 2000,
Carroll et al. (2003) concur that the programs for preparing pre-service teachers are not adequate in terms of developing skills and closing the gap between what is known about inclusive practice and what happens in inclusive classrooms. According to Carroll et al. (2003), Dempsey (2012a), and Selzter-Kelly (2013), current programs frequently overemphasise knowledge acquisition to the detriment of equipping teachers with practical skills for teaching a diverse range of students with varying abilities.

The theoretical design rationale and content composition of pre-service courses and their impact on future inclusive practice is particularly important and requires further scrutiny (Billingsley et al., 2004; Darling-Hammond, 2005; Darling-Hammond et al., 2005; Fore et al., 2002). The literature related to PST education courses supported this view and identified a number of issues that contributed to the efficacy of PST preparation to address the challenges associated with inclusion. These include: the lack of a common theory base driving the design coherence of teacher education programs (Snyder, 2012); insufficient depth of preparation (Carroll et al., 2003; Forlin et al., 1999); the nature and knowledge needed by teachers in inclusive settings as well as how teachers acquire this knowledge (Florian, 2011, p. 319; Stephenson, O'Neill, & Carter, 2012); inadequate linkage of university classrooms to professional experiences (Koehnecke, 2000; Mitchell et al., 1997); and empirical evaluation of these factors. Many of these overlap with the benchmarks identified earlier in the literature regarding inclusive practice within schools.

There has been limited empirical research undertaken examining the effectiveness of different PST education programs. As noted in chapter one, a key problem seems to be that many teacher education programs keep repeating the same practices they criticise (Zundans-Fraser, 2014). For example, although many respondents to the Australian Quality Matters: Inquiry into Teacher Education (Ramsey, 2000) noted that empirically-based research was promoted as critical to the field and as a key way to identify and validate effective models of teacher education, little of this style of research is actually carried out (Burkhardt & Schoenfeld, 2003; Hempenstall, 2006; Ure et al, 2009).

There is also a lack of empirical research about how well these PST programs work. Grima-Farrell (2012) noted, in her extensive review of the research literature about
closing the research to practice gap between PST and school settings, that very few empirical intervention studies examined the effects of PSTs, even though this was seen to be an area of major importance. Of the studies described in her review, only three could be included as intervention research: Darling-Hammond (2000); Miller, George, and Fogt (2005); and Winn and Zundans (2004). Darling-Hammond (2000) utilised a large data set across policies, case studies, staffing surveys, and national assessment progress reports on school students, and found that were strong correlations between PST preparation and school student achievement in reading and mathematics. Miller et al. (2005) found that the organisational structure of schools using a teaming approach assisted research to practice translation for PSTs completing onsite professional development to supplement their coursework. Winn and Zundans (2004) reported on 40 PST teachers working with 20 school students who had difficulties with literacy and numeracy, and the close collaboration between schools and universities was seen to be of ultimate importance.

In these three instances (Darling-Hammond, 2000; Miller et al. 2005; Winn & Zundans, 2004), other themes that emerged from Grima-Farrell’s (2012) literature review included: the requisite responsiveness of university programs in ensuring the practical preparation of teachers; the need for collaboration; and the need for continued support. Feedback was also required at multiple levels within schools and universities. A responsive and coherent PST course structure contributed to a unified approach in preparing PSTs (p. 54). None of these studies continued their research into the settings where PSTs were working and exhibiting target requisite skills.

Ludlow, Mitescu, Pedulla, Cochran-Smith, Cannady, Enterline, and Chappe, (2010) proposed that the lack of empirical research has amplified the debate about who should take control of accountability in higher education, what kinds of evidence are appropriate, and how different forms of evidence should be used. There is a need for validation of teacher education programs as well as consideration of the issues identified by Florian (2011) and other researchers above. None of the studies described so far examined the extent to which what was taught in a PST program was learned and whether that learning translated to classroom practice.
2.5.1 Evaluating PST courses that teach inclusion.

Empirically evaluating and validating PST courses that teach research-based pedagogies, and investigating the links to professional experience, are of the highest importance (Dawkins et al., 2008; Szabo et al., 2002).

Cochran-Smith and Zeichner (2005) conducted a major study in the United States within teacher education to develop a substantial evidence base for practices used in pre-service teacher education. The major focus of this work was on PSTs in general, however, Cochran-Smith and Zeichner devoted a chapter to the requirements of inclusive education. The chapter noted five major elements that were required in a successful inclusive education PST preparation program: 1). Establish a shared language in the teacher education curriculum to support collaboration between special and general educators; 2). Establish collaborative clinical practices; 3). Ensure competence in pre-service teachers before graduation; 4). Support beginning teachers in the first three years after graduating; and finally 5). Promote shared governance that reflects collective responsibility between stakeholders in pre-service teacher education.

These elements resonated with the earlier discussion on requisite school indicators of successful inclusion. Cochran-Smith and Zeichner (2005) discussed a shared language to enable communication between curriculum areas of the education degree. A shared language in schools was seen to be essential for communication in collaborative groups solving programming issues for students. Ensuring PST competence related to the implementation fidelity of research-based pedagogies required to ensure school student progress (Walton & Rusznyak, 2013). Support for beginning teachers and shared governance would assist the maintenance and generalisability of these skills in the PSTs’ graduate experience (Costley, 2013). The combination of these collective elements would positively enhance the self-efficacy for PSTs. These requirements noted by Cochran-Smith and Zeichner (2005) were echoed in those noted by Florian (2011) six years later, that the knowledge required by inclusive teachers refers to the skills of using high fidelity research-based pedagogies and collaboration to work productively with colleagues.

The call by Florian (2011) to identify key challenges in reforming PST preparation includes support for beginning teachers in the first three years after graduating; and promoting shared governance and collective responsibility between stakeholders in
pre-service teacher education. This involves validating practices used in PST preparation to empirically determine if these inclusive education benchmarks are actually being achieved.

There have been several reviews of empirical studies on learning to teach (Sleeter, 2014; Wideen, Mayer-Smith, & Moon, 1998; Wilson, Floden, & Ferrini-Mundy, 2002). Of the 196 research articles from the top four Ranked Impact Journals published up to 2012, Sleeter (2014) found only 28 of them (14%) did something other than report or comment on existing research. Further, she found that little attention was given to teaching students with diverse needs. She refers to the almost complete absence of this attention to diversity as an example of the ‘mini-silos’ in teacher education research (i.e., within rather than across communities of practice of staff and faculties involved with PST education. p. 6).

At the time of Cochran-Smith’s (2005) review, studies of the preparation of PSTs to work with students who have disabilities were also sparse. Few studies have described the specific organisation and practices used in higher education that prepare PSTs to work with diverse populations of students (Jimenez-Silva & Olson, 2012). In general, there have been few empirical studies on learning to teach students with diverse needs and even fewer where direct observation of actual learned practice in applied settings has been the dependent measure. To determine if similar results are still occurring since the reviews of Cochran-Smith (2005) and Sleeter (2014), the following section will describe the method employed to review the more recent literature on inclusion in PST.

2.5.1.1 Method employed to search the literature on inclusion in pre-service teacher education.

To determine the status of recent research into the field of teacher education literature, a search was conducted utilising the Scopus database and employing the descriptors of “pre-service teacher education and inclusion” in title of article, abstract, or keywords. The purpose of the following table (Table 2.1) is to illustrate the scope of studies and their inclusive education intent. Rankings taken from Scopus are defined as: “SCImago Journal Rank (SJR2): a metric based on the idea that ‘all citations are not created equal’. It counts the number of times an article published in the previous three years is cited in the current year, divided by the total number of items published by the journal
during that period. Unlike the Impact Factor, the subject field, quality and reputation of the journal have a direct effect on the value or ‘weighting’ of a citation.”

**Table 2.1**

**Journal Rankings Using Descriptors “Pre-service Teacher Education and Inclusion”**

<table>
<thead>
<tr>
<th>Journal name</th>
<th>Scopus journal rankings</th>
<th>Total articles located</th>
<th>Articles after descriptors were applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal of Teacher Education</td>
<td>2.222</td>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td>Teaching and Teacher Education</td>
<td>1.378</td>
<td>305</td>
<td>0</td>
</tr>
<tr>
<td>International Journal of Special Education</td>
<td>1.278</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>Journal of Education for Teaching</td>
<td>0.923</td>
<td>196</td>
<td>2</td>
</tr>
<tr>
<td>European Journal of Special Needs Education</td>
<td>0.785</td>
<td>74</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>795</td>
<td>9</td>
</tr>
</tbody>
</table>

*Note: only the 5/17 top ranked are included here. The 17 results returned included 20 articles after descriptors were applied.*

Only 0.01% of articles published returned results for “pre-service teacher education and inclusion”. The original articles were confirmed with others located via searches using Web of Science and Ulrichsweb using the same descriptors and resulting in a further 10 articles making a total of 30. A more extensive search was conducted through EBSCOhost (Education) as it included the most complete selection of databases including: Academic Search Complete, SocINDEX, Psychology and Behavioral Sciences Collection, ERIC, Education Research Complete.

As a means of focusing on recent research, studies were included in this review if they were published in a peer reviewed journal, if they were scholarly articles, if they were published between the years of 2000-2014, and if they included a focus on inclusion in pre-service teacher education contexts. Descriptors were introduced in the following sequence: pre-service teacher education (124,605 references); inclusion (3,293...
references); research-based including similes such as evidence-based, validated, rigorous (35 references). Terms such as ‘integration’ and ‘mainstreaming’ were not used as these were seen to be older terms in the inclusion literature, referring to an earlier idea of integration). These 35 articles and the 30 identified in Scopus were then analysed further as many overlapped (i.e. were the same as others identified). An abstract search using the same descriptors identified a final 30 articles that specifically presented detailed discussion of inclusion within PST courses. The five omitted mentioned PST in relation to inclusion but focused on a much broader definition of inclusion such as inclusion of gifted students or of students with cultural differences. Details of the included articles are presented below.

Procedures and practices that ensure successful inclusion and outcomes have been documented in the earlier sections of the chapter (for example: the use of research-based pedagogies like explicit teaching, the use of sophisticated language for problem solving and positive attitudes). The purpose of Table 2.2 is to summarise these 30 studies that described teacher preparation for inclusion in pre-service teacher education. The table will illustrate:

- Indicators or focus areas of inclusion that were being investigated;
- The methodology being employed;
- The dependent variable investigated;
- How the dependent variable was investigated; and
- Whether implementation results into practice were measured (i.e., whether there was efficacy data of the pre-service teacher preparation program).

Table 2.2 further elaborates on the search to determine if dependent variables associated with successful inclusion were included in empirical studies and whether they were directly or indirectly measured. Studies were considered empirical if quantitative or qualitative methods were employed and results were reported. They were not considered empirical if they involved a discussion or commentary about issues related to inclusion. Where a construct was measured by self-report then a self-report approach was viewed as a direct measure. Where a construct was enacted and could be observed, a self-report would be seen as indirect. A direct data collection source involved data collection that directly measured the variable of interest (for example, a survey to measure self-efficacy would be a direct measure of reported self-efficacy). An indirect measure does not directly measure the variable of interest (for
example, use of a survey to measure the quality of teaching when that implementation could be observed directly in a classroom setting (Blank, 2013; Bronkhorst et al., 2013; Cook & Tankersley, 2012).

The studies have been organised into the following sections and following the table will be unpacked under the numbered headings:

- Validation of instruments to measure inclusive education (2 studies)
- Evaluation of specific impacts on inclusive education (Self-efficacy) (13 studies)
- Evaluation of specific inclusive education skills (The use of pattern language for collaborative problem solving) (2 studies)
- Evaluation of specific inclusive education knowledge (Pedagogical content knowledge) (4 studies)
- Evaluating inclusive education skills in simulated settings (3 studies)
- Evaluating inclusive education skills in practice (5 studies)
- Knowledge about inclusive course design and evaluation (5 studies)
<table>
<thead>
<tr>
<th>Article</th>
<th>Target area</th>
<th>Research methodology</th>
<th>Dependent variable</th>
<th>Direct or indirect measurement of DV</th>
<th>Measures effect of PST reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block, Hutzler, Barak, &amp; Klavina (2013)</td>
<td>Self-efficacy in PST specialising in inclusion and physical education.</td>
<td>Empirical, exploratory and confirmatory factor analyses to validate a self-efficacy (SE) instrument using 486 physical education teachers.</td>
<td>Self-efficacy</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Zundans-Fraser &amp; Lancaster (2012)</td>
<td>Self-efficacy and the impact on inclusion of course designed using embedded design.</td>
<td>Empirical confirmatory factor analyses using 41 PSTs.</td>
<td>Self-efficacy</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Bruder, Dunst, Wilson, &amp; Stayton (2013)</td>
<td>Early childhood PST perceived self-efficacy beliefs about family-centered practices, teaming practices,</td>
<td>Empirical. ( N = 2,287 )</td>
<td>Self-efficacy</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Study Authors and Year</td>
<td>Research Questions</td>
<td>Methodology</td>
<td>Outcomes</td>
<td>Direct/Indirect</td>
<td>Action</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------</td>
<td>-------------</td>
<td>----------</td>
<td>----------------</td>
<td>--------</td>
</tr>
<tr>
<td>Study</td>
<td>Research Question</td>
<td>Methodology</td>
<td>Variable</td>
<td>Type</td>
<td>Finding</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>------------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Hamman et al. (2013)</td>
<td>Self-efficacy and the impact of practicum in an inclusive setting.</td>
<td>Empirical, structural equation modelling.</td>
<td>Self-efficacy</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$N = 337$.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lancaster &amp; Bain (2007)</td>
<td>Self-efficacy about inclusion based on type of field experience undertaken.</td>
<td>Empirical, survey scale completed by $N = 125$ PSTs in three different settings.</td>
<td>Self-efficacy</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Lancaster &amp; Bain (2010)</td>
<td>Self-efficacy about inclusion comparative study.</td>
<td>Empirical survey scale completed by $N = 30$ matched sample participants from each condition.</td>
<td>Self-efficacy</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Malak (2013)</td>
<td>Attitudes and acceptance of inclusion in Bangladesh.</td>
<td>Empirical, one-on-one interviews were conducted with $N = 20$ pre-service teachers.</td>
<td>Attitudes</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Symeonidou &amp; Phtiaka (2014)</td>
<td>Training requirements for inclusion for Greek-Cypriot teachers'.</td>
<td>Non-empirical, discussion.</td>
<td>Not Applicable</td>
<td>Not Applicable</td>
<td>No</td>
</tr>
<tr>
<td>Varcoe &amp; Boyle (2013)</td>
<td>Attitudes towards inclusive education.</td>
<td>Empirical, survey scale completed by participants who completed the elective unit on inclusion compared</td>
<td>Attitudes</td>
<td>Direct</td>
<td>No</td>
</tr>
</tbody>
</table>
to those who did not. *N* = 193.

<table>
<thead>
<tr>
<th>Article</th>
<th><strong>Target area</strong></th>
<th>Research methodology</th>
<th>Dependent variable</th>
<th>Direct or indirect measurement of DV</th>
<th>Measures effect of PST reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bain et al. (2009)</td>
<td><strong>Use of pattern language</strong></td>
<td>Empirical, uninterrupted time series design with <em>N</em> = 54 participants.</td>
<td>Frequency and sophistication of pattern language used when writing reflections</td>
<td>Direct</td>
<td>Yes</td>
</tr>
<tr>
<td>Arthur-Kelly, Sutherland, Lyons, Macfarlane, &amp; Foreman (2013)</td>
<td><strong>Pedagogical content knowledge</strong></td>
<td>Non-empirical discussion of changes in programs across Australia and New Zealand.</td>
<td>Not Applicable</td>
<td>Not Applicable</td>
<td>No</td>
</tr>
<tr>
<td>Article</td>
<td>Target area</td>
<td>Research methodology</td>
<td>Dependent variable</td>
<td>Direct or indirect measurement of DV</td>
<td>Measures effect of PST reported</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Jin, Yun, &amp; Wegis (2013)</td>
<td>Differentiating curriculum to enhance inclusion.</td>
<td>Non empirical, discussion.</td>
<td>Not Applicable</td>
<td>No</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>Bain et al. (2009)</td>
<td>The effect of differing teaching conditions on PST mastery of pedagogical content knowledge.</td>
<td>Empirical, counterbalanced groups for different teaching conditions. $N = 90$.</td>
<td>Pedagogical content knowledge about inclusive pedagogies</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Florian &amp; Rouse (2009)</td>
<td>Curricular reform in PST course using Schulman’s concept of three apprenticeships: the head (knowledge), the hand (skill and doing), and the heart (attitudes and beliefs) about inclusion.</td>
<td>Empirical, qualitative mixed methodology coding the lectures and tutorials delivered by staff. $N = not available$.</td>
<td>Inclusive education delivery of course content by staff</td>
<td>Direct observations of lectures</td>
<td>No</td>
</tr>
<tr>
<td>Fox-Turnbull &amp; Snape (2011)</td>
<td>Description of cognitive apprentices. Inclusion situated within three constructivist theories.</td>
<td>Empirical, mixed methodology with data collected from reflective quotes, direct observations and interviews. $N = 30$.</td>
<td>CoP functioning to produce a final assessment product linking theory and scaffold practices involved.</td>
<td>Direct observations of CoPs and indirect sources such as analysis of quotes in reflections and assignments, interviews.</td>
<td>No</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>---</td>
</tr>
<tr>
<td>Leaman &amp; Flanagan (2013)</td>
<td>Teacher education methodology as a way of bridging the theory to practice gap.</td>
<td>Empirical, discussion of role play as situated learning (RPSL) and naturalistic inquiry self-study of PST. $N = 35$.</td>
<td>Use of higher order thinking in a meta-commentary about a lesson.</td>
<td>Direct</td>
<td>Yes, of meta language</td>
</tr>
<tr>
<td>Article</td>
<td><strong>Target area</strong></td>
<td>Research methodology</td>
<td>Dependent variable</td>
<td>Direct or indirect measurement of DV</td>
<td>Measures effect of PST reported</td>
</tr>
<tr>
<td>In practice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>---</td>
</tr>
<tr>
<td>Klein, Taylor, Onore, Strom, &amp; Abrams (2013)</td>
<td>Scaffolding PSTs with yearlong inclusive internship in New Jersey.</td>
<td>Non-empirical. Creating a partnership with schools for an apprenticeship model of PST. Based on theory of creating a third space for PST.</td>
<td>Not Applicable</td>
<td>Not Applicable</td>
<td>No</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Article</th>
<th><strong>Target area</strong></th>
<th>Research methodology</th>
<th>Dependent variable</th>
<th>Direct or indirect measurement of DV</th>
<th>Measures effect of PST reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynch &amp; Smith (2012)</td>
<td>Translating theory and research into praxis.</td>
<td>Empirical, grounded theory and Case Study and analysis of final year capstone projects. T-tests and ANOVA statistical tests used on questionnaires. $N = 221$.</td>
<td>Graduate experience, course satisfaction</td>
<td>Indirect, questionnaires completed by practicum task mentors. And also graduates from the pre-existing and new programs.</td>
<td>No</td>
</tr>
<tr>
<td>Moran (2014)</td>
<td>Links to practicum.</td>
<td>Empirical, results from two years of surveys ($N = 262$) in Australia.</td>
<td>Links to Practicum</td>
<td>Indirect</td>
<td>No</td>
</tr>
<tr>
<td>Sela &amp; Harel (2012)</td>
<td>Introducing AR so teacher educators can teach PSTs how to conduct AR as a means to investigate their own practice in inclusive settings.</td>
<td>Empirical, Action Research and grounded theory to analyse reflective texts. $N = 30$.</td>
<td>Use of action research in classrooms</td>
<td>Indirect, document analysis of reflective writing component.</td>
<td>No</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Title</td>
<td>Methods</td>
<td>Empirical, grounded theory and Case Study analysis of final year capstone projects. T-tests and ANOVA statistical tests used on questionnaires. $N = 221$.</td>
<td>Graduate experience, course satisfaction</td>
<td>Indirect, questionnaires completed by practicum task mentors. And also graduates from the pre-existing and new programs.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Lynch &amp; Smith (2012)</td>
<td>Translating theory and research into praxis.</td>
<td>Empirical, grounded theory and Case Study analysis of final year capstone projects. T-tests and ANOVA statistical tests used on questionnaires. $N = 221$.</td>
<td>Graduate experience, course satisfaction</td>
<td>Indirect, questionnaires completed by practicum task mentors. And also graduates from the pre-existing and new programs.</td>
<td>No</td>
</tr>
<tr>
<td>Pugach &amp; Blanton (2012)</td>
<td>Analysis of content in dual certificate programs to prepare PST for inclusion.</td>
<td>Empirical, content analysis of program descriptions, course syllabi, and related program documents. $N = \text{not available.}$</td>
<td>Content of PST course content</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Snyder (2012)</td>
<td>Analysing adult PST learners in terms of Mezirow’s transformative learning. 10 steps of Mezirow’s transformative learning framework.</td>
<td>Empirical, qualitative case study of 4 students: archived data, participant generated data and face to face interviews. $N = 4$.</td>
<td>Stages of PST learners as they progress through their course</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Stephenson et al. (2012)</td>
<td>Web-based analysis of PST content used to teach about inclusion.</td>
<td>Empirical. $N = \text{not available.}$</td>
<td>PST course content</td>
<td>Direct</td>
<td>No</td>
</tr>
<tr>
<td>Young &amp; Florian (2013)</td>
<td>Using a methodological memo (MM) in a research and development project designed to reform initial teacher education (ITE) to Reporting on course development</td>
<td>Empirical, qualitative, mixed methods in a study of the theoretical and pedagogic processes</td>
<td>Reporting on course development</td>
<td>Direct</td>
<td>No</td>
</tr>
</tbody>
</table>
interrogate theory/practice divide. involved in developing a new ITE course. \( N = \text{not available} \).
Table 2.2 demonstrated that researchers were in fact answering the calls for more thorough investigation into inclusive education courses and of indicators known to make a positive difference to students with diverse needs. Discussion of the articles will proceed based upon the way they were organised in the table:

- 2.5.2 Validation of instruments used to measure inclusive education (2 studies)
- 2.5.3 Evaluation of specific impacts on inclusive education (Self-efficacy) (13 studies)
- 2.5.4 Evaluation of specific inclusive education skills (The use of pattern language for collaborative problem solving) (2 studies)
- 2.5.5 Evaluation of specific inclusive education knowledge (Pedagogical content knowledge) (4 studies)
- 2.5.6 Evaluating inclusive education skills in simulated settings (3 studies)
- 2.5.7 Evaluating inclusive education skills in practice (5 studies)
- 2.5.8 Knowledge about inclusive course design and evaluation (5 studies)

2.5.2 Validation of instruments used to measure inclusive education.
Both studies in this section were empirical and used direct measurement of the dependent variable: the construct of self-efficacy. Attitudes and perceptions of inclusion were studied by Block et al. (2013). Using statistical methods, the researchers measured the self-efficacy of physical education PSTs when facing challenges including teaching students with intellectual disabilities. The scale referring to intellectual disabilities (ID) resulted in a high Cronbach's alpha reliability (0.86) and confirmatory factor analysis demonstrated good goodness-of-fit, thus, the content and construct validity of the instrument was supported.

The self-efficacy questionnaire (SEIPD) used in this thesis was evaluated in a study by Zundans-Fraser and Lancaster in 2012. This validation found two interpretable factors of interest: ‘personal teaching efficacy (PTE)’, and a ‘skill level’ factor. The PTE factor accounted for 48.12% of the item variance and the skill level factor accounted for 10.78% of the item variance. Results from this study were similar to results found by Lee, Patterson, and Vega (2011) and concurred with Palmer (2006) in confirming the existence of other mediating factors, apart from those delineated by Bandura (1997) that may impact on the self-efficacy experienced by pre-service teachers.

These studies involved validation of the measures used for the self-efficacy construct, however they did not measure subsequent skill development or disposition. The
following sections discuss more specific skills, knowledges and indicators that enhance inclusion.

Figure 2.6 below highlights the focus of each section being covered within the pre-service teacher education arena. This Figure indicates the fact that all inclusive education skills, knowledges and attitudes are more likely to be achieved if a course is based on sound theoretical design.

![Figure 2.6](image)

*Figure 2.6. The integration and evaluation of skills and attitudes in PTE courses and the potential of theoretical design for PTE courses.*

2.5.3 Evaluation of specific impacts on inclusive education (self-efficacy).

As described earlier, personal self-efficacy has the potential to be a strong predictor of whether effective practices will be implemented by teachers (Al-Awidi & Alghazo, 2012; Avery & Meyer, 2012; Lastrapes & Negishi, 2012). When teachers have a high level of self-efficacy concerning effective pedagogies to use with students who have learning difficulties, they have been shown to persist longer with these students and evaluate student progress more frequently to the benefit of their students (Holzberger et al., 2013; Woodcock, Hemmings, & Kay, 2012).

A total of 13 research articles were located that evaluated the impact of PST self-efficacy on inclusive education. Two articles were discussions and 11 articles were empirical and included direct measures of the self-efficacy construct. The existing body of work on teacher education programs related to inclusion remains dominated by measures of PST self-efficacy and attitudes towards inclusion. The importance of positive attitudes towards inclusion was discussed by Agbenyega and Klibthong (2012). Burnout relating to negative attitudes and lower self-efficacy was discussed by
Symeonidou and Phtiaka (2014). While these papers were informative, they were not empirical and did not involve measurement of a dependent variable relating to the PST indicator of focus. Agbenyega and Klibthong, and Symeonidou and Phtiaka provided examples of the way some institutions tackled the issue of inclusive education and research to practice gaps, as well as providing rich discussion as a starting point for further empirical investigation.

Eleven of the empirical articles in Table 2.2 directly measured changes in PST self-efficacy using direct measures of the construct. Bruder et al. (2013) made comparisons between the self-efficacy of pre-service and in-service early childhood teachers and results indicated that greater self-efficacy was experienced following pre-service program completion. While this is a direct measure of self-efficacy using a questionnaire for pre/post data collection, this study did not directly measure the actual impact on the focus areas noted (family-centred practices, teaming practices, assessment, IFSPs/IEPs, and instructional practices of the participants).

In a study of self-efficacy related to PST experiences, Lancaster and Bain (2007) compared mentoring of high school students, classroom support for primary students, and in-class lectures and tutorials as the control for efficacy building experiences. They found statistically significant differences between pre and post occasions for all three different conditions. However, there were no significant differences between the groups. The greatest gains in self-efficacy were experienced by the cohort that did not involve an applied experience. The findings from this study suggested that the assumed link between the amount of exposure to students with disabilities and the attitudes towards teaching these same students could not be made. The amount of exposure to students with disabilities during pre-service training did not necessarily covary with more positive attitudes amongst PSTs. The authors suggested that a more detailed analysis of the types of course experience was required to establish the maximal impact on PST self-efficacy.

A follow-up study by Lancaster and Bain (2010) sought to extend the original research through a more focused comparison between a classroom-only condition using a specific educational design approach and the combination of a classroom condition with an additional direct applied experience in an inclusive setting. Again, the results indicated that the inclusion of an applied experience in the course did not necessarily
covary with greater gains in self-efficacy. In this study, the greatest gain was made by the group that experienced a particular course design approach that did not include an applied experience, although the differences between the two groups were not statistically significant.

Gedžune and Gedžune (2013) provided a snapshot of a means to engage PSTs in reflection and investigated the impact of this on their attitudes towards inclusion. When discussing e-learning (that is learning using on-line environments) as a means to enhance inclusion, the authors utilised discourse analysis to identify 15 codes for terms PSTs used to describe underlying assumptions about inclusion that they held. The authors suggested that the framework uncovered could be used by future teachers to help them make sense of inclusion. While e-learning created an environment that could be utilised by PSTs, this study did not make the connection to determine if PSTs actually improved in any of the skills associated with inclusion.

Hamman et al. (2013) used a quantitative approach to determine the effectiveness of associate teacher collaboration with PST students whilst on practicum. PSTs completed questionnaires about their collaboration with cooperating teachers, the focus on inclusive instruction in their practicum setting, and their efficacy for providing inclusive instruction. Results from structural equation modelling indicated that PST self-efficacy was significantly influenced by the amount of collaboration including sophisticated discussions they had with their supervising teachers about students who had diverse learning needs.

Interest in self-efficacy was noted by authors internationally. For example: the concern about attitudes towards the principle of inclusion was investigated by the following authors: Cologon (2012) in Australia; Haq and Mundia (2012) in Brunei; in Bangladesh, Malak (2013) reported unfavourable attitudes and uncertainties of PSTs; Varcoe and Boyle (2013) focused on attitudes as a mediating variable for self-efficacy; and likewise, perceptions of PSTs were studied in Spain by Chiner and Cardona (2013); and, finally, Eriks-Brophy and Whittingham (2013). These authors made self-efficacy comparisons between primary, early childhood and secondary PSTs.

Two of the articles in Table 2.2 were non-empirical and provided insights into attitudes towards inclusive education and training requirements (Agbenyega &
Klibthong, 2012; Symeonidou & Phtiaka, 2014). The remaining 11 studies were empirical and included direct measures of the self-efficacy construct (Bruder et al., 2013; Chiner & Cardona, 2013; Cologon, 2012; Eriks-Brophy & Whittingham, 2013; Gedžune, & Gedžune, 2013; Hamman et al., 2013; Haq & Mundia, 2012; Lancaster & Bain, 2007; Lancaster & Bain, 2010; Malak, 2013; Varcoe & Boyle, 2013). The articles described here support the view that PSTs require positive attitudes and perceptions about the inclusive environments they find themselves in so that they will persist longer and evaluate student progress more frequently to the benefit of their students (Holzberger et al., 2013; Woodcock et al., 2012).

However, attitudes and self-efficacy alone constitute a limited definition of what is required for teacher effectiveness in their future practice in inclusive settings. None of the 13 articles noted in Table 2.2 investigated self-efficacy and actual practice in the same study. Further investigation was needed into the skills and knowledge required for inclusion. Mastery of these skills and knowledge were found to be mediating factors that lead to higher levels of self-efficacy for PSTs. One of these mediating skills is the ability to communicate with others using specific and sophisticated pattern language.

2.5.4 Evaluation of specific inclusive education skills: Use of pattern language.

Use of a common professional pattern language in collaboration with colleagues is important to inclusive educators as a means to solve problems around students, their needs and learning at a sophisticated level. A shared pattern language aids the ability to provide feedback to each other about teaching approaches, student need and decision making (Friend & Cook, 2013; Hennessey & Dionigi, 2013).

Two articles were located that identified the use of pattern language for collaborative problem solving as skills that enhance inclusion. Firstly, Hamilton-Jones and Vail (2013) used a qualitative case study ($N = 12$) conducted over a semester of coursework through document analysis and interviewing ($N = 5$), to build a better understanding of PSTs’ perceptions and beliefs about collaborating with general educator partners in school settings. The focus was on collaboration in schools, but rather than directly observing the collaboration process itself, data collection involved indirect means such as document analysis and interview about the processes that the PSTs followed. Participants reported that after completing coursework they felt prepared to collaborate
and confident in their ability to do so in the future. Reflecting about collaboration resulted in PSTs being more aware of the collaboration around them, and more able to critically assess collaborative work. These findings supported coursework in collaboration as a mechanism to prepare PSTs for the collaborative aspects of their profession. Knowledge of the collaborative process steps were evident, however, the skills in actual implementation (for example: sophisticated use of pattern language) were not directly assessed (Hamilton-Jones & Vail, 2013).

Bain, Lancaster, and Zundans (2009) investigated whether the use of professional lexicon (pattern language) was differentially effected in terms of frequency and also sophistication following collaborative group meetings with peers during tutorial times. Data were taken pre and post occasions and analysed using parametric statistics. Results indicated that statistically significant improvement in use of terms followed each teaching topic and sophistication of use also increased. The use of collaborative processes was not reported, however, direct measures of the pattern language terms were utilised in this instance.

Consistent pattern language used for collaborative problem solving is cited as requisite for successful inclusion (Hamilton-Jones & Vail, 2013; Hennessey & Dionigi, 2013). Pedagogical content knowledge of research-based pedagogies would inform the pattern language being used in collaborative meetings.

2.5.5 Evaluation of specific inclusive education knowledge: Pedagogical content knowledge (PCK).

Pedagogical Content Knowledge (PCK) (Shulman, 1987) is essential if research-based inclusive practices are to be implemented with high degrees of fidelity. Four articles were located within the current search: three were commentary and discussion papers, and one was empirical utilising a direct measure of PCK.

Pedagogical content knowledge, evidence-based instructional knowledge, authentic mentoring networks and professional development planning were discussed and framed within a comparison between programs in Australia and New Zealand by Arthur-Kelly et al. (2013). The authors reviewed international trends and key policy and legislation platforms in both New Zealand and Australia (for example AITSL, 2015) in order to explore some of the challenges in enhancing PST programs to take
on more inclusive perspectives and content. Using document analysis between two PST programs, the authors found the following facilitators for enhancing PST programs: embedding policy and practice in national and international contexts; integration of course structures and content with the broader education faculty; listening to the views of practicing teachers; exploring PSTs’ values, beliefs, attitudes, knowledge and concerns about engaging with and teaching students with diverse learning needs; building evidence-based programming and instructional knowledge, skills and practices for effective differentiation and adjustments; contributing to authentic mentoring networks; and encouraging professional development planning.

The article provided an extensive commentary and suggestions with no direct measures of the efficacy of the design included.

As a potential means for closing the research to practice gap between PST programs and settings, Dooly and Sadler (2013) discussed two groups of PSTs in Spain and the United States who were required to work together in an online environment to give each other peer feedback and evaluation of pedagogy implementation. Multimodal data was collected over two years of the tele-collaborative project during the collaboration between the two groups of PSTs. Using tools such as Moodle, Skype, emails, wikis, Second Life and podcasting, the students collaborated with each other to provide feedback on activities and teaching sequences. Data analysis included meeting transcripts, audio files, and final teaching portfolios. Using ethnographic methods of analysis, results indicated that the online collaboration enhanced teacher PCK development through opportunities unavailable in more traditional teacher education classrooms. It also enabled PSTs to better make connections between theory and practice through implementation fidelity.

Theory as a means of driving course design and closing the research-to-practice gap was discussed by Jin et al. (2013). The authors proposed a curriculum framework (PCK) based on the Transtheoretical Model to train pre-service physical educators for inclusive education. Three curriculum stages (lecture-focused, lectures with teaching practicum, and internship-focused) and the stage-matched strategies adapted from the Transtheoretical Model were discussed as promising ideas to systematically restructure pedagogical content to effectively promote inclusion. Again, no direct measures of the efficacy of the design were included.
Bain, Lancaster, Zundans and Parkes (2009) included direct measures of gains made in PST content knowledge acquisition and compared three different learning methods to determine which method was the most effective for PSTs in their learning of PCK. The learning methods included self-study, peer tutoring, and structured collaborative learning groups. Collaborative groups using a cooperative learning structure resulted in the largest effect size of $d = 0.3$. The collaborative learning effect size supports the methods that require learners to actually share and elaborate on their PCK. This was the only study to incorporate a direct measure of effect size.

Research calls for high degrees of PCK for the pedagogies known to have a positive impact on the learning of students with additional needs (Cook & Tankersley, 2012; Kretlow & Helf, 2013). The current literature provides evidence of a research-to-practice gap regarding PCK (Antil et al., 1998; Grima-Farrell, 2012; Grima-Farrell et al., 2011; Hennessey & Dionigi, 2013; Jenkins et al., 2003). This was a disconcerting finding given the benefits of these practices are based upon research studies where their procedural features were implemented with high degrees of integrity (Slavin & Lake, 2009), as well as the scant research that directly measured PST pedagogical content knowledge.

PST programs have attempted to solve this research-to-practice gap by incorporating simulated settings into PST learning contexts. As well as distinct skills and knowledge, the current literature search identified studies that used direct measures of inclusive education skills, knowledge and dispositions for the data collection that occurred in simulated settings.

2.5.6 Evaluating inclusive education skills in simulated settings.

Three articles were located that investigated inclusive education skills in simulated settings. These settings involved preparing role plays and simulated exemplars for PSTs to engage with. The three articles illustrated a means of measuring PST skills in simulated settings, they were all empirical, and they all used direct measures to collect data.

Florian, Young, and Rouse (2010) described ways that the structure and content of an initial teacher education program for primary and secondary teachers was revised to ensure that social and educational inclusion was addressed within the core program.
Schulman’s (2005) conceptualisation of professional learning provided the framework for investigating all of the variables required to prepare inclusive educators: apprentices of the head (knowledge), hand (skill and doing), and heart (attitudes and beliefs) (in Florian & Rouse, 2009). The direct measures mentioned here focused on the lecture content being delivered to determine whether the PSTs could identify if any praxis existing within Schulman’s framework. The article did not include measures of the impact of this structure on actual inclusive education skills of PSTs.

Fox-Turnbull and Snape (2011) presented an example of an empirical study where PST pedagogy was analysed using direct observations of the PST community of practices (CoP) to complete a final assessment product. They described three constructivist theories: situated cognitive apprenticeship, modelling and scaffolding from Vygotsky’s Zone of Proximal Development. These cognitive apprenticeships were enacted through processes used by CoP where participants were able to discuss links to theoretical frameworks in their reflections. A mixed methodology was employed to analyse the data collected from reflective quotes taken from assignment work and then triangulated with direct observations and interviews. Results indicated that PSTs were able to delineate scaffolding that was afforded through use of CoP processes, while the direct measure of steps in the collaboration processes were not analysed. Authors also included indirect measures to complement the data sources, such as quotes from reflections and interviews with PSTs.

In a study by Leaman and Flanagan (2013), lecturers used role-play to model pedagogies for PSTs. Direct observations of three-hour videos were used to elicit PST self-study. Students were able to call for a halt in the role-play to question the rationale for particular actions being used at any given point in time. Qualitative results were presented in the form of quotes to illustrate higher order thinking and metacognitive understanding after analysing the lessons. The direct measure of PSTs in this study was the use of metacognitive language used to describe pedagogies that the PSTs witnessed, it was a direct measure of their language used to describe the pedagogies. In other words, Leaman and Flanagan (2013) measured meta-language describing research-based pedagogies, but not the actual use of those pedagogies. This study is reported here to illustrate the use of simulated settings in PST programs.
Florian and Rouse (2009) discussed ways that PSTs were able to accurately link content in simulated form from lectures to the model provided by Schulman (2005, in Florian & Rouse, 2009). There remains a huge metacognitive leap to generalisation of actual skills in practice. Fox-Turnbull and Snape (2011) also demonstrated PST ability to link a theoretical background to the process of collaboration, however there was no direct observation of PST practice in collaborative groups. Leaman and Flanagan (2013) directly measured the pattern language being used by PSTs in their simulated setting.

A further step used by PST programs to reduce the research-to-practice gap involved the practice of inclusive skills in-situ. The following section presents articles that used a combination of direct and indirect measures for the data collection when interventions occurred in practice.

2.5.7 Evaluating inclusive education skills in practice.

Five articles from Table 2.2 evaluated inclusive education skills in practice, usually while PSTs were involved in mandated practicum experiences. These occurred in regular classes in schools while PSTs were under the supervision of qualified classroom teachers. One article was non-empirical (Klein et al., 2013), while the remaining were empirical. Klein et al. (2013) discussed the scaffolding employed in a year-long internship program as a means of enhancing partnerships between universities and schools. By creating an apprenticeship model of PSTs with schools, the authors explain a theory of creating a third space for PST, a space away from the usual lecture and tutorial format and also more supported than the usual practicum placements. They described structures put in place to create a new space for inclusive education. Areas addressed included: epistemology, curriculum and clinical practice; and incorporated roles for K-12 teachers, university staff and mentoring partnerships. The authors provided a thorough description of course key concepts and assessment items.

Of the remaining four articles described in this section that were empirical, Lamb et al. (2013) were the only authors who used a direct measure of their dependent variable (that being PST reflection). Lamb et al. (2013) determined ways to enhance and measure reflective practice of PSTs whilst on their school practicum placement. In their study, 23 participants were assigned roles of being, as well as having, a buddy for
lesson observation. They completed training in being a peer reviewer and used a template of questions in order to assess videoed lessons. Results discussed reflective abilities that developed as a result of the process. Presentation of these data was not provided and only PST quotes about the process were reported in the results section. The study does, however, describe ways to enhance the reflective abilities of PSTs, and the notion of a peer observer could be used for direct measurement of inclusive education indicators in future investigations.

Using grounded theory, Lynch and Smith (2012) compared two groups of participants in order to understand final year curriculum products. They described the curriculum products developed whilst on practicum in terms of grounded theory principles. The authors used statistical methods followed by semi-structured interviews to analyse questionnaire data, and found that graduates of the new program rated their expectations as having been met at a higher level than those of the pre-existing graduates. The new graduates also perceived that they were better prepared than their predecessors. The details of what changed in their teacher preparation to produce this result was not elaborated. The mentors in schools found no statistical differences between the two groups. Lynch and Smith (2012) did not directly measure PST skills when translating grounded theory into PST practice.

In many teacher education programs, professional experiences were included as a means of enhancing PST understanding about teaching and the profession. The challenge the programs faced were the lack of places available in schools and, at times, the unpredictable quality of the placements given that some teachers are time-poor, are not always good models of effective teaching practice, or they lack the skills of articulating their practice. Moran (2014) provided an empirical study that described two years of surveys to seek findings that investigated links to practicum. Results from two years of surveys (N = 262) suggested that while the program did not negate all the difficulties associated with ensuring quality placements, it did provide some solutions that assisted in improving the professional experiences of PSTs. The authors did not measure directly the actual practice of PSTs, but relied on survey information from PSTs and their placement mentors.

Sela and Harel (2012) employed an action research approach to determine if such a process would have a positive effect on PST strategy implementation. Rather than
directly observe the steps of action research being carried out, they relied on document analysis and reflective writing of PSTs about the process of using action research. The authors used action research as a tool to determine if it was useful to enhance reflection of practice. The 30 participants consisted of three groups for comparison: practicing teachers, those studying a graduate level program, and PSTs in the final year of their undergraduate program. Use of action research was seen to assist new teachers to become knowledge creators instead of just knowledge consumers. Data revealed discussion of feelings about using action research as a process rather than observing its effects in practice. Levels of reflection by the PSTs were considered as technical, practical or of a critical standard. Translation of these reflections into practice were not measured in a direct way as they had been by Bain, Lancaster and Zundans (2009) using SOLO.

The studies described in this section did not measure actual changes of the PSTs’ skills, knowledge and practices. Rather, they focused on indirect measures of issues other than those indicators in the literature that relate to successful inclusion (Lynch & Smith, 2012; Moran, 2014; Sela & Harel, 2012). Lamb et al. (2013) used direct measures, however, these measures did not determine whether a PST course actually worked. They did not determine if there was actually any change in behaviour as a result of knowledge and skill development in real settings. The key words in the abstract by Lamb et al. (2013) included a recognised inclusive education indicator, however, this was measured indirectly through PST reflection.

As well as the individual attempts to close the research-to-practice gap for PST inclusive practices, overall course design has been an additional means investigated to enhance inclusive education. Actual course design to enhance inclusion and evaluation of the course is taken up in the following section.

2.5.8 PST course design and evaluation.

There were five empirical articles located that investigated course design as a means to enhance inclusive education. The study by Lynch and Smith (2012) investigated a university program design where school-based practicum mentors completed questionnaires. A comparative study was designed involving mentors in schools who had supervised only PSTs from the pre-existing program; those who had supervised PSTs from the new program; and those who supervised PSTs in both. Mentors were
also asked to comment on their PSTs in terms of their skills and attitudes towards the inclusive classroom. Lynch and Smith (2012) used indirect methods to study the variables of interest via use of graduate experience questionnaires to compare the skills of the groups of participants and consequently their course. They found that graduates of the new program rated their degree higher than did the existing graduate PSTs. The majority of mentors from both programs considered preparing teachers to be a joint responsibility and that the new framework achieved this end in a better manner than the original program. While results from the questionnaire and semi-structured interviews provided results in favour of the newly devised course, specific details of the course were not included in the study.

Based on a content analysis of program descriptions, course syllabi and related program documents, Pugach and Blanton (2012) examined the curricula of three fully merged teacher education programs that were redesigned to prepare teachers for the full range of diversity in their student populations. In these programs, graduates could earn a general and special education elementary license simultaneously. Results suggested that attention to disability was more prevalent than attention to other social identity markers such as race, class, culture or language. While direct measures were employed to determine course content, these did not translate into direct measures of changes in PST skills. They did not study whether students acquired these skills in practice.

Snyder (2012) examined four in-depth case studies of PSTs as they progressed through their course. Using multiple survey methods of data collection, the findings were mapped onto a theoretical framework. The findings indicated five emergent themes: the benefit of a spiral curriculum; effective writing discourse; authentic learning; experiential learning; and collegial relationships. The idea of effective writing discourse relates to the use of a professional language to discuss programming and planning as part of the teacher’s daily life when acting in collegial relationships. Authentic learning was seen to occur as real emergent feedback was given in a continuous fashion to assist future actions as well as understand those occurring in the past. Snyder (2012) provided a direct measure of Mezirow’s phases of learning that the PSTs pass through during their study. The study did not describe evaluation of the course based on actual PST skills.
Even though Stephenson et al. (2012) discussed PST course content and its effect on the self-efficacy of beginning teachers, it is included in the discussion here as the focus from the abstract was on the evaluation of a course. This study reported final-year Australian PSTs’ sense of efficacy scores and the sources of information that contributed to it. Results showed that beginning teachers had a good sense of efficacy, and classroom management was not differentiated from instructional or engagement efficacy. Further, personal qualities and physiological and affective states predicted self-efficacy scores, and having a number of opportunities to practice skills was associated with a higher sense of efficacy. The indicator of self-efficacy was measured directly and suggestions were presented to provide more practice for PSTs, no other skills relating to the course were measured.

Young and Florian (2013) described how a methodological memo (MM) was used to reform initial teacher education. The MM was used to explore what often remains an opaque part of the research process the relationships between theory and practice that underpins many research studies. Yet, without detailed accounts of the research process provided by the authors, it remained difficult to understand this relationship. In this case, a memo was used to record the methodological and analytical dilemmas that were encountered in a study of the theoretical and pedagogic processes involved in developing a new course so that primary and secondary teachers were better prepared to respond to increasing diversity in schools. The authors suggested that developing and sharing the research process transparently might help clarify contested concepts and provide the foundation for richer theoretical debate. It remained a discussion and no measures of the impact of the course development were evident. In terms of attribution, that is, the extent to which there was measurement of what is taught and how well it was implemented in order to link the findings on the dependent measure to the intended curriculum or practice in PST, none of the studies included measures of implementation or processes that would make such attribution possible.

In summing up, of the articles relevant to PST course design and evaluation, only one of the empirical studies used an indirect measure of the dependent variable (Lynch & Smith, 2012). The remaining four empirical studies used direct measures of the intended dependent variable: the content of PST courses (Pugach & Blanton, 2012) and the links between courses and theoretical frameworks (Snyder, 2012; Young & Florian, 2013). With the exception of Stephenson et al. (2012), who investigated the
impact on PST self-efficacy in relation to course content, these dependent variables did not match the skills, knowledges and dispositions noted earlier in the chapter. None of the studies showed that the PSTs learned the skills in class first and then translated them to the practice setting.

2.5.9 Summary.
As Cochran-Smith had discussed in 2005, there are still very few examples of empirical studies that focused on recognised indicators of inclusion and used direct measures of the dependent variable to determine if changes occurred for the PST participants involved. Self-efficacy was the only exception, where 33% of studies located used this variable as the sole focus of their study. In addition to studies being carried out on individual indicators of inclusion, a number of studies were located that specifically discussed ways to unlock the processes of PST course design and evaluation; however they did not include direct measures of PST capabilities (Pugach & Blanton, 2012; Snyder, 2012; Young & Florian, 2013). Few empirical studies discussed specific organisation and practices used to support PSTs in inclusive education. Compared to the findings reported by Cochran-Smith and Zeichner (2005), this current search located nearly twice the number of empirical studies compared to discussion or commentary articles.

Table 2.2 demonstrated that research is, in fact, answering the calls for more empirical analysis of PST programs in relation to inclusion. Of the studies described in Table 2.2, 22 utilised empirical methodologies, one-third of the empirical studies utilised surveys to document PST experiences through pre/post program surveys, and three utilised implementation observations. Eight incorporated mixed methodologies and two were quantitative.

The empirical studies primarily employed participant-completed survey-based data or qualitative research methods, and many used self-study of courses or programs at the institute where they were taught. These were often rich sources of data. It is important to also develop an objective empirical understanding of the constellation of factors that impact on PSTs’ sustained use of evidence-based effective pedagogies. In the past, researchers have used observational data to investigate the integrity and efficacy of pedagogy implementation (Huberman & Miles, 1984, in Gersten, Baker, & Lloyd, 2000). In this current search, only two used observational methods: one measured
lecturers’ use of modelled pedagogies (Florian & Rouse, 2009) and the other involved analysis of video recorded lessons (Lamb et al., 2013).

None of the studies included measures of implementation or process that would make attribution possible. That is, the extent to which there was measurement of what was taught and how well it was implemented in order to link the findings on the dependent measure to the intended curriculum or practice in PST courses. While these more recent studies responded to the need for more objective data, there remained a critical need to investigate the indicators collectively that research tells us impact on PST knowledge and skills (Cochran-Smith & Zeichner, 2010). Only in instances where PST outcome data on more indicators of inclusion are actually measured can the design of PST education courses be empirically evaluated and validated as a total package.

No study was located that showed the effect of PSTs on correlates of requisite skills required for inclusive education (i.e., pedagogical content knowledge, professional language about research-based pedagogies, self-efficacy) and actual PST behaviour in inclusive classrooms (Florian & Rouse, 2009; Jordan, Schwartz, & McGhie-Richmond, 2009; Stephenson et al., 2012; Van Laarhoven, Munk, Lynch, Bosma, & Rouse, 2007). There remains a need for studies that determine not only the efficacy of a particular aspect or feature of a PST course, but also to study the extent to which that learning can be translated into classroom practice. Given the relationship among pattern language, pedagogical expertise and self-efficacy described throughout this review, a need still exists to undertake research that shows whether PSTs can actually develop those skills as expressed in the work undertaken by graduate teachers in inclusive settings.

In addition to thoroughly researching the commonly identified indicators essential for successful inclusion, validated theoretical frameworks are still required to design PST education programs to provide a template for the maintenance and sustainability of skills acquired. Theoretical frameworks would enable consistency of reliable validated skills and practices once in place. Theoretical design of pre-service teacher education courses will be discussed in the following section.
Figure 2.7. Theoretical course design that enhances inclusive skills and attitudes.

2.6 Theoretical Design behind Pre-service Teacher Education

A factor often blamed for the inadequacies of PST programs is the lack of adequate theory to drive practice and also to enable the translation of theory into practice. This is a focus of this study where a theoretical design principle was applied to the design of a course in inclusive education.

The consequences of the theory/practice divide in PST education continue to be exacerbated by an increasing expectation that schools will become more responsive to the learning needs of an increasingly diverse population of students. Table 2.2 and the related discussion showed there were a number of empirical as well as commentary or discussion papers that linked a theoretical perspective to practice. Both empirical and commentary studies all fell into two categories: studies that situate PST learning within an existing theoretical framework (Section 2.6.1); as opposed to those that use theory to design a course (Section 2.6.2). These will now be elaborated further.

2.6.1 Studies that situate PST learning within an existing theoretical framework.

Three of the articles discussed possible links to theory by examining theoretical frameworks and linking current practice to the theoretical structure. Dooly and Sadler (2013) describe strategies where PSTs were encouraged to use collaborative learning through computer-mediated communication where collaborative learning is achieved using computers rather than face to face collaboration in order to discuss appropriate teaching pedagogies. The authors frame their discussion around Socio-constructivism and situated learning. Klein et al. (2013) scaffolded PSTs with a year-long internship
in New Jersey. By creating a partnership with schools for an apprenticeship model of PST, they explained the support as a theory of creating a *third space* for PSTs, a space away from the usual lecture and tutorial format and also more supported than the usual practicum placements. Leaman and Flanagan (2013) discussed *role play as situated learning (RPSL) and naturalistic inquiry self-study of PSTs.*

A further seven authors reported on empirical studies that aimed to link theory to practice. Fox-Turnbull and Snape (2011) described constructivist theories: situated *cognitive apprenticeship, modelling and scaffolding from Vygotsky’s Zone of Proximal Development (ZPD).* These cognitive apprenticeships were enacted through processes of CoP. A mixed methodology study was conducted with data collected from reflective quotes taken from assignment work and then triangulated with direct observations and interviews. Examples are presented of the higher level thinking obtained by the PSTs as they participate in this collaborative and cooperative exercise and reflected on their learning. The data showed that while undertaking a constructivist based activity, PSTs gained a rich understanding of theory related to technological practice and they made relevant links to learning in the classroom. Gedžune and Gedžune (2013) used action research and e-learning to enhance inclusion. Discourse analysis was employed to analyse reflective journal of PSTs. Lamb et al. (2013) used self and peer reflection, focused group interviews, and analysis of videoed lessons whilst students were on practicum. Grounded theory methodology informed an emergent model of reflective practice. Lynch and Smith (2012) also used grounded theory to understand final year curriculum products produced by PSTs. Sela and Harel (2012) combined grounded theory to analyse reflective texts in conjunction with action research as a means for PSTs to investigate their own practice. Snyder (2012) analysed PST learners in terms of Mezirow’s 10 steps of *transformative learning framework.* Finally, Young and Florian (2013) discussed use of a *methodological memo* (MM) to interrogate the theory/practice divide and to study the theoretical and pedagogic processes involved in developing a new PST course. These authors appear to have been using theory as a means of data analysis.

These authors were analysing and situating PST learning within an existing theoretical framework and investigating whether PSTs could employ particular theories in tasks related to PST coursework assessments. All of these studies used theory in terms of *student experience* rather than as a framework to design courses. Use of theory is
called for that is more proactive in nature in order to deal with issues related to learning and behaviour in inclusive classrooms as they arise. The three studies that incorporated grounded theory to inform an emergent model (Lamb et al. 2013; Lynch & Smith, 2012; Sela & Harel, 2012) did not include any follow-up validation of the links to the theoretical model. The focus remained at the level of discussion and description without measuring the actual impact on practice. Theory needed to provide for the possibility of empirical feedback in order to adapt to the possibility of change as required.

The literature discussed consequences of loose connections between theory and practice, including a critical lack of feedback measures and processes (Anderson & Freebody, 2012; Husebo, 2012) and ambiguous implementation (Muncey & McQuillan, 1996). The usual suspects for these loose connections were noted in terms of supports that are lacking including: time, money, leadership, resources, teacher attitude and professional development (De Valenzuela, Connery, & Musanti, 2000; Sarason, 1996); however, these suspects were not interrogated any further than being noted.

In 2012, Husebo discussed a ‘Meta–Approach’ combining Action Research with Interpretive Cultural Approach. She wanted to determine if embedding a ‘research process’ such as action research into PST practicum would enhance the ability to use knowledge and skill in practice. Action research was embedded in a CoP with school associate teachers and PSTs on practicum. Meetings were audio and video taped. By working with the school staff as a CoP with the PSTs, connections were made between PSTs and their learning connections as well as progress toward solving problems. Rather than focus on a single theory, Husebo (2012) described use of a meta-approach to bridge the course knowledge with a particular pedagogical approach using action research in a CoP (Kemmis, 2010). She studied student teachers on internship and discussed methods of teaching religious education. By involving school teachers in the research, it ensured the PSTs were not left on their own to try to navigate problem solving in real settings, and created a bridge between theoretical and practical issues. Collaborating with school-based teachers meant that mutually dependent and complementary communities were developed to help PSTs bridge the theory and practice gap. Jimenez-Silva and Olsen (2012) also referred to the incorporation of a CoP (a teacher-learner community) in teacher preparation courses. Husebo (2012)
concluded that the cycles of action research used with any theoretical framework would effectively provide feedback in proactive as well as reactive nature.

In summary, further studies are still required to measure both the baseline and longitudinal outcomes of such an organisation of learners. One persistent problem in teacher education is that teachers continue to teach as they were taught in schools, which makes change difficult to implement and even harder to sustain (Anderson & Freebody, 2012). The construct of CoP is grounded in sociocultural theories of learning where participation and collaboration with others enables human social and cognitive development to occur. CoP in a professional context requires a professional language to discuss enactment of professional practice.

While the use of CoP does have an impact on course content and the manner in which the content is taught, the guidelines for course design and development are not specific in the literature. Action research has sound empirical basis (Gedžune, 2014; Waterman et al., 2010), as does the use of case study research (Sela & Harel, 2012). The results from Jimenez-Silva and Olsen (2012) were based on PST reflections, anonymous course evaluations, and semi structured interviews. No specific guidelines were offered—just the use of the Action Research framework, and no objective data was taken on PST results of their learning—just their self-reported opinion of the same. Most of the studies used theory as a term of reference for an aspect of a course. As well as studies that situate PST learning within a theoretical framework, other studies focused on theory used to design courses for PSTs.

2.6.2 Theory being used to design courses.
Over time, authors have discussed various means of using theory in PST education courses. Some, for example, suggest that theory should provide the basis for conceptual frameworks for course development (Francis, 2002; Roth, Lawless, & Tobin, 2000; VanderVen, 2000). VanderVen (2000) proposes that Lilian Kratz’s work with Chaos theory could provide a schema for the integration of theory and practice however specifics were not offered as to how this would actually could be translated into practice.

As well as theoretical frameworks being the driver, other authors emphasised ‘process solutions’ for pre-service course design. For example, some simply called for
additional time in the field as a means to make explicit links between theory and university courses (Koehnecke, 2000). The implication of this is that linking theory to practice will happen by osmosis simply by PSTs being in-situ.

These theories and process solutions were espoused as being innovative; however, they were not subjected to empirical scrutiny regarding their efficacy for PSTs. Given issues associated with bridging the theory to practice gap for PSTs, it could be expected that more sophisticated approaches to negotiating a solution should be determined. Evidence-based methodology is called for.

Heydon (2005) argued that some practices in inclusive education are too positivist and linear in nature and suggested that each theory and set of practices from the field of inclusive education had something to teach us, but that not all should necessarily be employed. Heydon suggested one theory or set of inherent practices should not claim dominance over another and described an ‘ethical praxis’ of pedagogies to overcome the theory and practice divide. On closer analysis, the fifth tenet Heydon proposed resonated closely with the work of Bain (2007) where she described ‘ad-hocracy’ as an organisational structure that relied on innovation to cope with change. It is a structure where labour is shared and mutually agreed, decisions are made, and goals set by the collective stakeholders. Heydon’s tenets remained at the level of description rather than being developed into a tested and validated process.

Odom and Wolery (2003) concurred with Heydon (2005) and described the potential for a unified theory of practice to emerge from the early intervention field. They described tenets of the theory and supported these with evidence-based practices, which was an important addition to the literature as the use of evidence-based practice was not always mentioned, let alone substantiated. Rather than relying on a set of practices based on single perspectives, these authors were in agreement with Heydon that the theory and practices needed to form a ‘praxis’. They refer to this as a ‘theory of change’ based on the fact that there are shared exemplars among practitioners and scholars, each having a body of research as the foundation. These suggestions have progressed the discussion and description level of analysis but fall short of fully tested and validated processes.
While the above provide a platform for discussion, *testing of the theory being used to design courses* was not evident in most of the articles reviewed. In contrast, Bain (2007) posits a meta-theory and design process to enact these practices and to ensure the system employing them would be capable of dealing with changes in environmental issues on an ongoing basis. The *practical meta-theory* has been seen as necessary to fill the space between intent and action in inclusive education (Bain, 2007; Heydon, 2005; Odom & Wolery, 2003). Bain’s meta-theory was one that was evidence-based, addressed issues of scale, and attempted to account for change and adaptation to differing and continuing contextual circumstances as a means to exert greater influence over student learning. Bain’s (2007) approach was concerned with school reform and was embedded within the parameters of Complexity and Self-Organisation.

2.6.3 *Use of a meta-theory from the field.*

The following discussion will (i) explain the ways adaptive systems operate and incorporate self-organisation; (ii) situate human systems within complex adaptive systems; and (iii) situate complexity within education and higher education arenas. A discussion of school reform research and the detailed application of complexity and self-organisation is then provided.

Complexity theory has emerged over the last four decades based on the study of chaotic, nonlinear and dynamic behaviour of systems. Complexity theory focuses on *how* systems adapt, and how they self-organise when plunged into the realm of the unexpected (Bain, 2007). Complexity theory “complements and extends notions of traditional science by explaining the behaviour of systems that are emergent and constantly changing” (Brodnick & Kraft, 1997, p.3; Waldrop, 1993).

Complex systems operate via their agents, which are the individuals that exist within the systems. The agents operate simultaneously and the systems are characterised by these multiple interactions (Brodnick & Kraft, 1997; Stacey, 1995). Even though the interactive nature of systems leads to complex behaviour that can be difficult to understand, predict and manage, Kauffman (1995, p. 8) suggested that “profound order is being discovered in large, complex, and apparently random systems”.

Examples are often furnished from the natural environment, such as the research of Gordon (cited in Davis & Sumara, 2006, p. 19), who studied the lifecycle of ants and
related this lifecycle to a complex system that could be sustained over time. Robinson (2013) noted that complexity could well be related to human systems as well as those found in the natural environment.

Because complex systems are dynamic, unpredictable, spontaneous, adaptive and alive (Davis & Sumara, 1997; Fels, 2004; Urry, 2005; Waldrop, 1993), they are particularly important as an appropriate methodological framework in the examination of human relationships (Waldrop, 1993). The richness of interactions amongst agents allows the system to undergo spontaneous self-organisation. As Burnes (2005, p. 79) suggested, “there is no overall blueprint or external determinant of how a system develops; instead, the pattern of behaviour of the system evolves or emerges from the local interaction of the agents within it”. It is this self-organising ability that allows systems to adapt to their environment or particular context in order to survive and to maintain coherence throughout changing circumstances (Davis & Sumara, 2006, p. 455). It is necessary to consider how the parts of a system inform the collective--and the benefit of this type of continuing shift in education is that complexity “aims to understand the whole by understanding the interaction of its parts” (Phelps, 2002, p. 168).

The practice of education occurs within a complex system that includes individual agents, collectives of individuals, communities and larger cultural contexts (Davis & Sumara, 1997). As a result of advances in the study of complex systems, a need has been identified to revise and further examine approaches to the way in which human systems function, particularly in connection to educational programs (Buell & Cassidy, 2001; Burgess, 2004).

In education, complexity theorists such as Gough (1999) and Reason and Goodwin (1999), suggested that the value of complexity lies in its capacity to encourage “new ways of seeing and new language to describe such relationships and structures” (Phelps, 2002, p. 172). It is a system created for a particular need that definitely relies on networks at a human level, requiring agents to be interactive and connected through a relationship established by engagement with the system. Complexity theory “offers a different conceptual framework for the study of conceptual change” (Houchin & MacLean, 2005, p. 164) by providing the theoretical framework for combining alternative viewpoints into a coherent perspective on process (Stacey, 1995).
There are six common theoretical principles derived from the field of self-organisation and complex adaptive systems described in the model developed by Bain (2007) and employed by Zundans-Fraser (2014). These principles help to explain how systems can adapt to the ever-changing circumstances they face with diverse students populations, from a bottom-up intervention rather than top-down. Those principles are: simple rules, embedded design, emergent feedback, similarity at scale, dispersed control, and the common schema (Bain, 2007). These principles will be unpacked in a subsequent section (See section 2.6.4.2). When used together, these principles generate what is known as a self-generating capability that makes it possible for the parts of a system to work together so that the system can make itself (Urry, 2005) and continue to adapt to change. The following section will further unpack the empirical work that has been carried out using these principles of self-organisation within education systems.

2.6.4 The self-organising school.

Using a theory deeply embedded in complexity theory, complex adaptive systems and self-organisation within a school system, Bain (2007) provided an example of the large-scale application of the principles from complexity and self-organisation. He conducted a longitudinal field-based study over 11 years and found ways to deliver feedback to teachers as they delivered lessons using research-based pedagogies. The integrity of implementation fidelity for explicit teaching and cooperative learning was judged by teachers, their peers and students, and data were entered into a central database so that emergent feedback could be provided and future planning could be managed. The extensive program of research incorporated 1,600 direct classroom observations of key teaching approaches from 1997-2001, an eight-year study of student achievement, and more than 12,560 student evaluations of teacher implementation practices (Bain, 2007, p. 248). Results showed that implementation integrity in the self-organising school project resulted in research-based practice being sustained at scale over time. The school community was able to build a collaborative culture using consistent pattern language to describe their work. Bain (2007) demonstrated that theory can make a practical connection between ideas, systems and practice in schools, and that this connection can be maintained over time.
Additional support for this design was found through the Hong Kong school project (Bain et al., 2011). The project also built around the six principles of self-organisation and targeted capacity-building in a co-educational secondary school, initiated in 2009. There were four strategic goals for the project: (i) to integrate the principles of the self-organising school with the school’s vision and goals; (ii) to apply the principles to building school capacity in mastery teaching, cooperative learning, differentiated instruction, and the use of technology; (iii) to develop and use a lesson plan database to help build a school repository; and (iv) to build a system within the school for recognising and rewarding teachers. At the time of publication (2011), results from the project were progressing towards these goals, with explicit acknowledgement of the importance of leadership, clear vision and goals, and the need to create conditions where this type of innovation can occur.

While the text by Bain (2007) and the Hong Kong school project (Bain et al. 2011) both focused on school reform, the same barriers and issues are relevant to the higher education arena if we are to have a scalable effect when PSTs begin their work in the next phase of their professional lives, in the field of inclusive education. The following section investigates this notion of scale-up from university course work to teachers’ work in schools through the work of Grima-Farrell (2012, 2013).

2.6.4.1 Use of the theoretically designed course to investigate the research to practice divide between university coursework and practice in school settings.

Grima-Farrell (2013) investigated the significant gap that remained between our accumulated knowledge about what can work in classrooms and the extent to which evidence-based practices are used in sustainable ways. This inability to bridge the research to practice gap has had an adverse effect on the progress of inclusion in schools and the ability of individual teachers to respond to the needs of all students. Grima-Farrell examined factors that both enabled and interfered with the successful translation of research to practice in education settings based on a case study of six participants who completed a postgraduate Masters of Inclusive Education degree. The degree had been designed at a university using the six principles of self-organisation that had been studied by Bain (2007), Bain et al. (2011), and Zundans-Fraser (2014). The participants involved in Grima-Farell’s study completed a project that was carried out in their own school settings. The effectiveness of embedding
principles and processes in the course content was highlighted by one of the students interviewed as part of Grima-Farrell’s doctoral work:

I think they embedded things right from the start. It was heavy going and we were coming from so many different worlds. We lived and breathed the course. It had a life of its own. It was the readings, the research, and the collaboration and then we would put it into practice at school. It was putting the research into practice without us even knowing or realizing it. (‘Diane’ in Grima-Farrell, 2013, p. 246)

Another student, Sam, synthesised the various aspects of the course design in one of her responses. She mentioned the way that aspects were modelled in the design, the embedded nature of the theoretical framework throughout the course, collaboration and capacity-building through her ability to transfer knowledge to practice:

By replicating what we had done in the course was the biggest feature. The course was modelled so well. The whole course was collaborative and was embodied into a theoretical framework. The other really big thing in that course was the whole professional language around inclusion. It is not superficial it was that go deeper [sic], it was actually going deeper and it is the total comprehension, the language, the living of the course. (Sam, in Grima-Farrell, 2012, p. 250)

Based on the student feedback after completing the project, participants attributed their capacity and the success of their projects in part to the design of the postgraduate program. The notion of a process of development built through continuous cycles aligned well with the feedback cycles also embedded throughout this process. While Grima-Farrell’s (2012) study was not a study of the theory per se, it did show the efficacy of the theoretically designed program in terms of bridging the research to practice gap in inclusive education.

2.6.4.2 Investigation of inclusive education course and program design

In 2014, Zundans-Fraser investigated inclusive education course design, utilising the same principles of self-organisation as Bain (2007) and Bain et al. (2011). She identified ways a theoretical course design could address four major areas of concern currently facing university programs: (1) The need for a theoretical basis of course design and reform, (2) the need to reduce the theory-to-practice gap, (3) the need to
utilise collaborative practice, and (4) the need to examine and acknowledge the impact of institutional practice on reform.

To address this concern for a theoretical base to inform course design, Zundans-Fraser (2014) identified and described complexity theory as a design tool. The study involved applying the design-based methodology to enact six principles of self-organisation to create an inclusive education program. The data collected sought to determine how the theoretical principles of self-organisation could inform the program design process. This included studying the nature of the interaction among members of a design team, the collaboration and feedback process, how the methodology evolved and adapted over time, and the way courses were designed to provide a coherent, interrelated experience for postgraduate students.

A three-phased design process was employed to create a course design approach within the higher education context that was practical, responsive to contextual needs, and reflective of contemporary standards in the field of inclusive education. This research originated from an ethnographic theoretical orientation adopting a participant researcher stance, and use of case study as the strategy of inquiry. The participants in this study were six members of the Inclusive Education Faculty, and the design process examined was a longitudinal study of a change initiative within an Australian regional university. Throughout the case study, data were collected using qualitative methods with triangulation occurring through use of multiple sources of data collection as well as across phases of the study.

Zundans-Fraser (2014, pp. 461-462) found:

- Theory can, and should be, a powerful driver for course and program design processes.
- Course and program design processes need to be supported by institutional practice.
- Organisational change to process and culture needs to occur so that collaborative working and use of feedback are part of the everyday mode of working within an institution.
- The areas of need consistently identified in the literature require broad range change and cannot be addressed in isolation.
• Recognition, capacity building and encouragement of staff to work in different ways are critical.

Zundans-Fraser (2014) proposed that this meta-theory may provide guidance on what was needed, as well as how to enact this in the changing circumstances of PST education classrooms. The resultant model from Grima-Farrell in 2013 provided an understanding of scale-up from the teacher education platform, as well as at graduate level in schools, when dealing with students who have varied teaching needs.

Given the above findings and previous success in school reform (Bain, 2007; Bain, et al., 2011), program design in inclusive education (Zundans-Fraser, 2014), and in generalising the impact of a university program to practice in the field (Grima-Farrell, 2013), a need existed to pursue a more experimental focus on specific key principles of the meta-theory that framed the basis of these studies.

The following section will elaborate early case studies that investigated the principle of embedded design (one of the six principles comprising self-organising theory and the focus of this study).

2.6.4.3 The self-organizing design principles unpacked.

The self-organizing approach involves applying principles derived from prior work on a theory of self-organizing schools (Bain, 2007) to teacher education course design. The self-organising schools theory derived from the study of self-organizing systems (Brodnick & Kraft, 1997; Davis & Sumara, 2006; Robinson, 2013; Waldrop, 1993) involves the application of six design principles that enable change from the ground up which include: simple rules, embedded design, similarity at scale, emergent feedback, dispersed control and use of a schema to guide action and accommodate change (Bain, 2007). Simple rules could be described as guiding principles such as: Inclusion is the design team’s term of reference, and content will be based upon the research standards of our field. Similarity at scale refers to common features to courses within a program with elements such as: Self questioning, peer mediation, authentic assessment, advance organization, and concept mapping into a research driven framework. A common educational design is embedded across all courses. Emergent feedback happens at multiple levels involving the design team, the program design and course design. Emergent feedback is also embedded within the course itself where feedback
is provided to students as they complete components of learning. Dispersed control refers to giving PSTs to have control over their learning. This occurs when collaborative groups are set up and peers give each other feedback as well as the guided practice from the lecturer. *Embedded design* is seen a central principle that embeds all others into the design of the course.

Bain and his colleagues (2007, 2009) used preliminary case studies to investigate the impact of one of the program design principles (the principle of *embedded design*) taken from his earlier work with the self-organising school, to determine the impact on inclusive education indicators for PSTs. The first study investigated the impact of the *embedded design* (ED) principle on the dependent measures known to have a positive impact on teacher skill when working in inclusive settings (Ainscow et al., 2004; Gillies & Ashman, 2000; Klingner et al., 1999; Ware, 2003). These measures include: pedagogical content knowledge (Bain, Lancaster, Zundans, & Parkes, 2009); utilising the same independent variable; use of common professional pattern language (Bain, Lancaster, & Zundans, 2009); and degree of self-efficacy (Lancaster & Bain, 2007). These variables were investigated as within-subject group designs. All results showed positive changes—and some with significant results.

In 2009, Bain, Lancaster, Zundans, and Parkes sought to establish the differential effects on achievement of embedding evidence-based practice in the design of an inclusive education teacher preparation course. *Embedded design* involved creating self-repeating patterns in the instructional design of an undergraduate course by expressing essential design features at multiple levels in the teaching and learning experience. A repeated measures counterbalanced design was employed in the study, with self-study as the control condition between tutorial groups of PSTs. The results indicated that PSTs attained a mastery level knowledge of the course content that covaried with the application of the *embedded design* principle. The authors found a statistically significant difference in PST achievement as a function of the teaching approach (cooperative learning, peer-assisted learning, or self-study) employed as part of the *embedded design* process. The score/percentage differentials across the three conditions were not large, but favoured both cooperative and peer-assisted approaches over the standard self-study (control) approach, and indicated medium effect sizes.

Professional pattern language was investigated as the dependent variable by Bain,
Lancaster, and Zundans (2009). In this case study, classroom communities of practice were employed as a learning context for PSTs to develop their pattern language and as a vehicle for applying the *embedded design* principle. The study also sought to establish whether increases in the frequency and sophistication of pattern language use increased as the pre-service course progressed through four teaching cycles and PSTs learned more about inclusive approaches. The results indicated that pattern language frequency and sophistication covaried with participation in the course, and increased over time. A repeated measures ANOVA indicated statistically significant differences in pattern language use over the four teaching cycles ($F (3, 141) = 49.59, p = .0001$). As the PSTs progressed through the course, they used the pattern language of inclusive practice with greater frequency every time a new teaching cycle (i.e., new content) was added. The sophistication of response across all four questions also increased as the teaching cycles and *embedded design* principle was implemented in a manner that covaried with an increase in the frequency of pattern language use.

In 2007, a study by Lancaster and Bain examined whether participation in a 13-week undergraduate inclusive education course covaried with an improvement in the self-efficacy of pre-service elementary education teachers. The intent was to determine whether self-efficacy was influenced differentially by the type of field-based placement experienced by students in the course compared to a cohort who experienced a course design based on *embedded design*. The results showed that an improvement in PST self-efficacy covaried with participation in the inclusive education course. The field-based placement did not differentially affect self-efficacy at a statistically significant level.

What was now called for was further investigation of design principles that involved comparative groups in order to join the dots across course design and PST practice. To join the dots, elements of pedagogy, pattern language and self-efficacy, both in terms of learning from the course and actual teaching practice, all need to be investigated.

The principle of *embedded design* represents one of the principles investigated by Bain (2007), Grima-Farrell (2012) and Zundans-Fraser (2014). It is not the intention here to look at this principle in isolation, but rather to describe it as the *central principle* in the design approach that encompasses all of the others in complexity and self-organisation.
This centrality of *embedded design* was explained in the following way: Of the six principles of the meta-theory described earlier, simple rules were articulated through the *embedded design* process that were then scaled up to all levels of work. Emergent feedback created the conditions for constant formative adaptation of the work and dispersed control to allow for constant refinement of content covered. It was this interaction that permitted a system to “make itself” by constantly adapting to its changing needs and circumstances (Urry, 2005, p. 10). For example, a PST using professional pattern language of cooperative learning would provide feedback about a design problem that would be understood by all members of the team. The common understanding was that the team could pool its collective intelligence to solve problems. Any solution could have an impact at scale because it could be deployed at all levels in the process or product of the course. The latter was important in that the theory was designed to create the conditions for constant change and adaptation in order to address issues associated with the way courses were refined (Edwards, Gilroy, & Hartley, 2002).

When viewed in this way, theory in this context was required to represent more than ideas and concepts; it needed to explain the way things happen as well as the loftier attributions usually attributed to the namesake (Bain, 2007; Kuhn, 1962). The *embedded design* principle allowed this to happen and drew the principles together to form a comprehensive problem-solving tool.

Successful complex systems exhibit self-repeating patterns or similarities within their organisational structure (Waldrop, 1993). *Embedded design* is the principle that creates these self-repeating patterns by expressing the simple rules in the design of a system and embedding those design features in all other features (Bain, 2007). *Embedded design* created a predictable relationship between learning and teaching that was necessary for feedback to be shared and understood by all involved. As such, *embedded design* is proposed to be the *pivotal design principle* in the self-organising theory.

Zundans-Fraser (2014) also described the *embedded design principle* as a primary influencer, as well as a key self-organisational principle to the process of program design. She noted that embedding particular design principles throughout the program
ensured cohesion in both content and delivery. Simple rules employed meant the collaboration was to be used as a tool embedded throughout the program and similar design structure to all courses comprising the program. This was echoed in her participant feedback:

I think the consistency in the Master’s [program] across all [courses], was replicated in our setting because everybody knew what they were going to expect. I think the collaborative approach through the Uni masters project, the feedback; the consistency was replicated in my setting (Zundans-Fraser, 2014, p. 408).

We were presented with a lot of the research across different [courses]...You know learning about collaboration, working as a team, what was going to be good for the whole school and a lot of the stuff I had learned through my university [program], I was able to use when we were devising that program and implementing it in the school (Zundans-Fraser, 2014, p. 408).

A recommendation for future research provided by Zundans-Fraser (2014) was to investigate the effectiveness of a program design process where institutional practice conditions were created that embedded capacity building and the design process within the institution. The complex nature of the intersecting and interrelated factors that influenced successful program design was a key finding of this study. However, this study involved the efficacy of a design process when designing a postgraduate program for inclusive education. It did not address the efficacy of the program design or the impact of the specific design principles on PST learning and teaching. The current study, in part, takes up this challenge.

2.7 Summary of Concerns with Pre-service Teacher Preparation and Implications for the Current Study

The efficacy of research-based pedagogies and curriculum differentiation is well established in the literature. The need remains for consistent implementation fidelity on an ongoing basis to ensure successful inclusion of students with varying learning needs in our educational settings (Archer & Hughes, 2010; Blank, 2013; Dettmer, Knackendoffel, & Thurston, 2013; Dooly & Sadler, 2013; Gedžune, 2014; Hennessey & Dionigi, 2013; Hunter, 2011; Pace, 2011; Rosenshine, 2012).

The successful inclusion of students with additional needs in regular classrooms remains a challenge for teachers when research-based pedagogies such as explicit
teaching and cooperative learning are not translated into practice with fidelity. A sound understanding of pedagogical content knowledge about research-based strategies will assist teachers in translating these pedagogies into practice. Teachers need to feel competent and capable with these inclusive pedagogies (Grima-Farrell, 2012; Hueng, 2006; Kretlow & Helf, 2013).

As well as the ability to deploy evidence-based pedagogies, sustain their use and implement them to scale with integrity, collaboration skills using consistent professional language are required to mediate successful inclusion. When these skills of using consistent professional language and pedagogical expertise are evident, it is expected that self-efficacy and competence of teachers working with students who have varying needs will increase. Teacher efficacy is related to teacher behaviours in the classroom and is said to effect teacher levels of effort, planning, organisation, persistence and resilience. Attitudes, beliefs and self-efficacy of teachers also need to be improved in order to work successfully with students who have different needs (Avery & Meyer, 2012).

All of these indicators require collective consideration so the optimum results may be obtained for PSTs’ inclusive practices. Schulman’s (2005) concepts of ‘heart, head and hands’ directly links these areas required for research on inclusion. The ‘heart’ refers to attitudes and PSTs’ self-efficacy about inclusion; the ‘head’ refers to thinking about inclusion and being able to discuss inclusion in a metacognitive way using professional language; and the ‘hands’ relate to practical skills of implementing inclusive pedagogies (in Florian et al., 2009).

Many studies provided discussions of these issues, but very few incorporated empirical evidence to enable attribution of the skills and knowledge to the preparation program. Few authors have provided systematic investigation of the theoretical design or content of PST courses and what PSTs learn from them. The relationship between theory and practice is seemingly an endless challenge for those in PST higher education (Husebo, 2012, p. 455). PSTs are preparing for a profession where course knowledge and knowledge about the didactics, pedagogy and education are inextricably linked and related within teaching practice. The aim of teacher education should be to try to mirror and validate this complex relationship.
In spite of the rhetoric surrounding translation of theory into practice for PSTs, and common calls for action to be taken, criticism of lack of progress in these areas remains. Theory is advocated as a means to drive the practices taught to improve PST outcomes; however, rarely is there any check for implementation integrity to ensure these outcomes are actually improving, or that testing of the theory is occurring in empirical ways. In keeping with these findings, Sindelar and Brownell (2001) concurred that the gaps described by Rosenshine in 1986 still exist today, nearly thirty years later. It would seem that the current PST education programs have not succeeded in closing the gap for their undergraduate teachers. The American National Council for Accreditation of Teacher Education (NCATE) in 2010 concluded that teacher education is fragmented with course matter, theory and pedagogical principles taught in isolation (Wiseman, 2012). They go so far as to recommend that “teacher education be turned upside down” (Seltzer-Kelly, 2013, p. 133).

Preparing teachers for inclusive education has not received sufficient attention from researchers (Florian et al., 2010). The studies included in Table 2.2 incorporated analysis of PST education--and the indicators noted as important for teachers (PCK, pedagogies, identity and attitudes) were also included (Sleeter, 2014). No data was identified that focused on PST outcomes in practice. In addition, the interrogation of a theoretical base for course design to aid integration of these required features is still required for inclusive education.

Given the above findings and success in school reform (Bain, 2007), this study provides an extension for the work in program design in inclusive education carried out by Zundans-Fraser (2014), and in generalising the impact of a university program to practice in the field (Grima-Farrell, 2012). It would seem time to pursue a more experimental focus on specific key principles of the meta-theory at the course level.

Based on these results from Bain (2007), Grima-Farrell (2012) and Zundans-Fraser (2012), the literature provided a starting point to determine if the theoretical design had an impact on PST level of PCK of research-based strategies. The level and sophistication of professional pattern language used was investigated in terms of the fidelity of feedback PSTs will be able to provide as future members of collaborative CoPs. The level of self-efficacy was investigated as attitudes are often noted as barriers when teachers face complex classroom situations. Jordan et al. (2009) have shown
how teachers’ beliefs about their responsibility for teaching students with additional needs are reflected in their practice. There is a relationship between what teachers think about the nature of ability to learn and teacher willingness to accept responsibility for teaching all students. In these earlier studies, the three variables (PCK, professional pattern language, and PST self-efficacy) extended earlier work and were investigated in a quazi-experimental manner (Bain, Lancaster, & Zundans, 2009; Bain, Lancaster, Zundans, et al., 2009; Lancaster & Bain, 2007, 2010). The final variable of interest for this study was the most pivotal translation of those pedagogies to scale, beyond a course within a 32 course program in which they are initially studied and into future practicum experiences of PST students. There were no studies that empirically tested the use of a theoretical framework to generalised skills of inclusion.

In conclusion, this review provided four important drivers for the present study. They are:

1. The need to ensure that inclusive education teachers possess the skills required for inclusive practice and the capacity to communicate that practice in professional ways as they collaborate with others.

2. At present, there are few studies of pre-service teacher education (PSTE) the focus on driver number one. Whether PSTE courses are teaching these skills to requisite levels of mastery is not known. Nor is there currently a body of literature showing the efficacy of PSTE in providing graduate teachers with those skills.

3. While theory can provide a focus and organising framework for the design of PSTE, the literature to connect the design of courses to a theory base is limited. This lack of theory inhibits efforts to determine the efficacy of PSTE given that lack of an anchoring framework or set of design principles that can drive efficacy research.

4. Given driver three, a need exists to scaffold efforts to determine the efficacy of theory driven PSTE courses by establishing whether their theoretical underpinnings can be linked to the development of the skills described in driver
one and whether those skills are reflected in the actual practice of graduate teachers.

This study stands at the confluence of the four drivers. The research seeks to determine whether the implementation of a theoretically derived course design covaries with the development of pedagogical skills, self-efficacy and pattern language of PSTs, and whether those competencies are expressed in actual classroom practice. As such, the study seeks to contribute to the need described throughout the review for a more robust base of evidence about teacher preparation for inclusion.

As a result of the literature review and drivers, the following questions were deemed relevant to this study. The research questions are presented for two studies that form part of the current investigation. The purpose of this research was to determine whether a course developed using the principle of embedded design covaried with an increase in the degree of PCK attained; an increase in self-efficacy the PSTs have for working with students who have difficulty learning; with the use of professional language and the sophistication of that language; and how well the pedagogies may be implemented when in the field.

The research questions described here reflect the need to provide a more detailed understanding of PSTE described in contemporary literature and the results of a series of preliminary studies described earlier in the literature review.

The overarching research questions investigated here are twofold:

**Study one:**
Does the application of the embedded design principle to the design of a PST inclusive education course result in improvements in:

- PST mastery of course content in the form of pedagogical content knowledge (PCK)?
- The frequency and sophistication of pattern language used by PSTs?
- PST self-efficacy?

**Study two:**
Does the application of the embedded design principle covary with improvements in:

- Design of classroom practice used by PSTs in the professional experience setting?
• Implementation of classroom pedagogy used by PSTs in the professional experience setting?
• Self-reflection about design and implementation of pedagogies used in the professional experience setting?
• The frequency and sophistication of pattern language used by PSTs when in the professional experience setting?
• PST self-efficacy?

The following chapter details the methodology used to conduct the research and investigate each of the research questions.
Chapter 3: Method

Given the tremendous diversity present in our schools today and the lack of evidence that research-based strategies are employed with the requisite characteristics necessary to ensure success for all students (Gersten, Chard, & Baker, 2000; Slavin et al., 1996; Stevens & Slavin, 1995a; Vaughn et al., 2003; Wolford et al., 2001), researchers have struggled to scale effective practices in the field (Ryder et al., 2006; Vaughn et al., 2003). They also call for more to be achieved at the pre-service level to try to bridge the research to practice gap (Gersten, Fuchs, Compton, Coyne, Greenwood, & Innocenti, 2005). The current research investigates the impact of a theoretical framework used as a means of bridging the gap between training, policy, and practice of inclusive education.

The two studies described here investigated whether participation in a course learning experience based on design-theory intended to improve the effectiveness of a pre-service teacher (PST) course in the area of inclusive education covaried with increases in pedagogical content knowledge, pattern language, self-efficacy and classroom practice.

This work was part of a longitudinal program of research that applied principles of self-organizing systems to the design of PST preparation in inclusive education. The self-organizing approach involves applying principles derived from prior work on a theory of self-organizing schools (Bain, 2007) to a teacher education course and program design. The study described here focused on a pivotal principle; that of embedded design. The broader theory and the research that supports it (Bain, 2007) has been discussed in the literature review (see section 2.6.4).

Specifically, this research involved the application of the embedded design principle (ED) which operates as a pivotal design principle of the self-organizing theory, and incorporated studies investigating the application of ED on variables that have been shown in the literature to have an impact on teacher skill, their attitude, and their ability to instantiate inclusive practices.
A conceptual model used to explain the studies is represented as follows:

*Figure 3.1. The application of embedded design throughout study 1 and 2.*

The first study investigated the impact of ED on the dependent measures known to have a positive impact on teacher skill when working in inclusive settings (Ainscow et al., 2004; Gillies & Ashman, 2000; Klingner et al., 1999; Ware, 2003). These measures included: pedagogical content knowledge, use of common professional pattern language, and degree of self-efficacy. These variables were investigated using a between-subject group design with comparisons made across two experimental conditions: embedded design and applied experience design. Embedded design was chosen as it is a key principle from the theory of self-organization and applied experience incorporates the use of in-situ experience in tandem with class contact times and often reported as a favourable design in the literature. The applied experience design is used here instead of a standard control condition (where no treatment is afforded) as it is difficult to determine what a standard course model would be. Prior research shows that the efficacy of an applied experience is beneficial for PSTs, and by using this comparison condition a more competitive or demanding test of the embedded design can be achieved. The control group received instruction which was seen as a more robust approach to compare the ED condition with an approach that has been identified in the literature as a source of benefit or efficacy by giving PSTs an experience in an applied setting as part of a course experience. A detailed description of both conditions is included in Section 3.9 that follows.

Utilising the same independent variable, the second study investigated the impact of the embedded design principle on the design and implementation fidelity of teaching practices that PSTs used while on practicum. The study sought to establish whether any gain made in classroom learning of PSTs would transfer into the quality of teaching in a professional experience setting. This research aimed to extend the use of evidence-based investigation into this arena (Cohen, Manion, & Morrison, 2007; Gersten et al., 2005; Shaughnessy, Zechmeister, & Zechmeister, 2009). Study two was a continuation of study one, and as such shared a number of common elements (e.g., the overall design, use of the SEIPD and pattern language measures). However,
differences in participants, procedures and dependent variables and measures called for separate descriptions of those elements. In order to avoid confusion, they will be described separately as study one (S1) and study two (S2). For each study, this chapter will:

- List research questions and hypotheses;
- Describe the research design employed to investigate the independent and dependent variables of interest;
- Discuss threats to validity of the method and how these were countered in the current research;
- Describe the procedures followed for study one and two, including the researcher’s access to participants, settings and ethical considerations;
- Describe levels of the independent variable for experimental and control groups;
- Describe the instrumentation process used to collect and code data and issues of validity and reliability of each; and finally
- Describe the data analysis procedures.

3.1 Study One

The research questions described here reflect the need to provide a more detailed understanding of PST education described in contemporary literature. The remainder of the chapter will instantiate the details of the research designs and methods chosen in order to investigate the research questions posed.

3.2 Research Questions

The overarching research questions investigated here are:

Does application of the embedded design principle to the design of a PST inclusive education course result in improvements in:

- The mastery of course content in the form of pedagogical content knowledge (PCK)?
- The frequency and sophistication of pattern language? and
- PST self-efficacy?

3.3 Hypotheses

Hypotheses are stated in the ‘research’ and ‘null’ hypotheses forms. The research format has a numerical subscript; the null format has a ‘0’ as subscript.
H₁ = There will be a statistically significant difference in the levels of:
   a) The mastery of course content in the form of pedagogical content knowledge (PCK)
   b) The frequency and sophistication of pattern language, and
   c) The level of self-efficacy between the PSTs who experienced application of the embedded design principle to the theoretical design of an inclusive education course as opposed to those who participate in an applied experience condition.

H₀₁ = There will be no statistically significant difference between intervention condition in the levels of:
   a) The mastery of course content in the form of pedagogical content knowledge (PCK)
   b) The frequency and sophistication of pattern language, and
   c) The level of self-efficacy between the PSTs who experience the application of embedded design principle to the theoretical design of the inclusive education course as opposed to those who participate in an applied experience condition.

The research design that delineates the finer details and processes used for Study One follows.

3.4 Research Design

Research design may be seen as a philosophical framework that holds fundamental assumptions about research processes (Lankshear & Knobel, 2004; Verma & Mallick, 1999). This framework influences the design as well as the procedures followed (Cohen et al., 2007). In turn, the framework influences the methods and more specific techniques of data collection and analysis procedures employed (Fraenkel & Wallen, 2006; Gay & Airasian, 2003; Shaughnessy et al., 2009).

The experimental research design is most appropriate for testing hypotheses about cause and effect relationships. Important characteristics of experimental design that require consideration include the following: comparison of groups; randomisation; and manipulation of the independent variable, data collection and analysis employed (Cook
& Campbell, 1979). Each of these characteristics were elaborated and investigated in terms of this research.

Usually an experiment involves two groups: the experimental group and the control used for comparison (Cohen et al., 2007). This research employed a non-equivalent group design where one group is compared to another, specifically, a single variable design with one manipulated independent variable. The experimental group received a ‘treatment’ and the control received a different or no treatment. To receive ‘no’ treatment is impossible in educational research, so perhaps a better term is ‘comparison group’ rather than control (Fraenkel & Wallen, 2006; Shaughnessy et al., 2009).

In the present research, the treatment condition or independent variable (use of the embedded design principle) was applied to participants in one experimental condition while those in the comparison condition were exposed to a different course design referred to as the applied experience design. The applied experience design involved in- situ interactions with school students to work on individually devised programs.

The experimental and comparison groups were pre-tested, administered the treatment and then post-tested. Employing a pre-test / post-test design allows the researcher to assess the effects of changes in the educational environment; make comparisons both before and after the intervention; and make conclusions about the intervention once the threats to validity have been taken into account.

The means of manipulating the independent variable across treatment groups, and also the random allocation of the participants to either experimental or comparison condition (Cohen et al., 2007; Gay & Airasian, 2003), was based on circumstance rather than the preferred randomisation as the two groups were already formed. Where random allocation of participants is not possible, the design of research is referred to as quasi-experimental, which is an important alternative to true experimentation in the social sciences (Shaughnessy et al., 2009). In the present study, both groups of participants were enrolled in the Bachelor of Education Primary program, a four year undergraduate teacher preparation program, situated on different campuses so it was impossible to randomly allocate them to differing treatments (the embedded design cohort and the applied experience cohort were assigned to treatment or comparison
groups). The mandatory inclusive education course that was the focus of the study was taught by different instructors whilst adhering to the core objectives set down by the course profile and outcomes specified in the program (See Table 3.8 for a summary of the differences between the cohorts).

A set of conditions or treatments (i.e., the independent variables) were assessed in terms of their effectiveness to engender change on the dependent measures (the outcomes). In this instance, the effect of *embedded design* was the independent variable, and was determined based on the outcomes or dependent measures of interest in inclusive education which include: PSTs’ levels of self-efficacy, PCK, sophistication and frequency of professional language, and also the implementation of strategies whilst on practicum.

Utilising the independent variable and the dependent variables across experimental and comparison groups, the design for study one is illustrated in Figure 3.2 to represent building PST students’ capacity for inclusion:

![Diagram](image)

**Figure 3.2.** S1 Non-equivalent control group design.

The *non-equivalent between subject group design* is used extensively in the area of special education settings in the field of practicing teachers (for example: Angeloska-Galevska, 2004, 2005; Bottge, Rueda, LaRoque, Serlin, & Kwon, 2007; Brown, 2005; Gersten et al., 2000; Hausstatter & Connolley, 2007; Odom et al., 2005) and also in higher education (Borko, Liston, & Whitcomb, 2007; Johnson, 2004; Nonis, 2008). Its applicability seems logical for PSTs preparing for inclusive education (Coladarci, 1982; Cornford, 2002; Gordon & Debus, 2002; Johnson, 2004).
Fraenkel and Wallen (2006) used the design term ‘Static-Group Pre-test/Post-test Design’ (p. 272) which is represented as follows:

Table 3.1

<table>
<thead>
<tr>
<th>The Static-Group Pre-test/Post-test Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment group</td>
</tr>
<tr>
<td>Control group</td>
</tr>
</tbody>
</table>

The dashed line refers to the two groups already formed. $X_1$ symbolises the experimental treatment (the *embedded design* independent variable). $X_2$ symbolises the *applied experience* independent variable. As described earlier, the control group received a comparison condition which was seen as a more robust approach to compare the *embedded design* condition with an approach that has been identified in the literature as a source of benefit or efficacy by giving PSTs an experience in an applied setting as part of a course experience. $O$ refers to the observations of both groups occurring at the same time, both pre-interventions as well as post. The subscripts of 1 and 2 denote the fact that the observations were the same in both settings in terms of time and content (Shaughnessy et al., 2009). It is important to note that $O$ also represents several different co-variables of interest in this study. The changes following application of the independent variable were measured through use of pre- and post-tests for each of the dependent variables.

Given the current focus on the need to increase rigorous and scientific research in educational settings (AITSL, 2015; Gersten et al., 2005), consideration needs to be given to the validity of the chosen group design as it is a direct function of the degree to which the internal and external variables are controlled. To make causal inference from quasi-experimental design, researchers need to meet the same basic requirements of experimental research. Special cautions need to be taken regarding threats to validity (Johnson & Christensen, 2008). A meta-analysis carried out by Heinsman and Shaddish in 1996 (cited in Asenio & Johnson, 2001; Johnson & Christensen, 2008, p. 332), compared effect size estimates from randomised designs with non-equivalent group designs to determine the extent to which similar results would be obtained from studies using these two designs. The results were a strong endorsement of the quasi-
experimental methods as they suggested both designs would yield about the same effect sizes.

It is often the case that the independent variable may be confounded by extraneous variables, in the categories of participant or environmental variables, which make it difficult to determine the unique effect of each. Randomisation of treatments or participants is the best single way to control for many extraneous variables; however, as noted above, it is not always possible (Burns, 2000; Lankshear & Knobel, 2004; Shaughnessy et al., 2009). Gersten et al. (2005) provide a table of considerations to determine if research may reliably be called evidence-based, the majority of which were present in the current study.

Table 3.2

<table>
<thead>
<tr>
<th>Essential Quality Indicators for Groups and Quasi-Experimental Research Proposals (Gersten et al., 2005, p. 151)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicators</td>
</tr>
<tr>
<td>Conceptualisation underlying the study:</td>
</tr>
<tr>
<td>1. Is a compelling case for the importance of the research made? Is the conceptualisation based on well-designed studies and does it reflect the scope of extant knowledge?</td>
</tr>
<tr>
<td>2. If an innovative approach is proposed, is it based on a sound conceptualisation formed from sound research?</td>
</tr>
<tr>
<td>3. Are the research questions appropriate and stated clearly for the purposes of this study? Are valid arguments supporting the nature of intervention in the comparison group(s) presented?</td>
</tr>
</tbody>
</table>

The purpose of experimental analysis is to determine the effect of an experimental treatment in contrast to a control or comparison condition. Usually this entails an analysis of the post-test scores of the two groups. Thus, to get a reasonable estimate of the treatment effects, researchers needed to ‘control’ for the effects of initial differences. This can range from establishing differences of each group at selection to test score outcomes.
Cook and Campbell (1979) take this notion of control further and assert it is essential to clarify what is meant by *equivalent* and *non-equivalent* when distinguishing between groups of participants. They suggest the term *equivalent* does not simply imply that the group means scores at pre-test were the same. It rather refers to a situation where increasing the size of the groups would still maintain the same group means. Thus, according to Cook and Campbell’s (1979) definition of equivalence, it refers to the equivalence of expected (population) values *not* equivalence of obtained (sample) values. Conversely, *nonequivalent* means that expected values of at least one characteristic of the groups are not equal, even in the absence of treatment. In order to understand the nature of a group’s non-equivalence, the selection process must be understood thoroughly. Having the same pre-test scores can be *suggestive* but not sufficient to establish the groups as equivalent; in fact it may even be misleading. Thus to understand the differences, it is imperative to understand the selection processes used to assign participants to groups as well as the pre-test scores. In this instance, the pre-test scores were taken on all dependent variables as well as a detailed description of the selection process provided to enable a clear understanding of the group compositions.

A number of statistical methods assist with the problem of selection differences, such as analysis of variance (ANOVA), analysis of covariance (ANCOVA), analysis of variance with blocking or matching, and analysis of variance with gain score analysis. Some of these allow more insight into the structure of the data than others. However, no statistical technique can be applied without some form of bias occurring, so the analysis needs to be carefully tailored to fit each research situation. Gersten et al. (2005, p. 150) assert that if a hypothesis can “withstand scrutiny by multiple methods, its credibility is enhanced greatly.” As such, a number of methods that follow were considered for the current research once the benefits and weaknesses of each were scrutinised.

ANOVA specifies three components of data to be determined at post-test: the grand mean (average response of all individuals), the treatment effect (the average value that the treatment added in a treatment group), and the error or residual (the effect of all other factors that may have impacted on the results). Lankshear and Knobel (2004) refer to this design as causal comparative as the association between two or more variables being studied, where at least one of the variables is categorical, but no
attempt is made to manipulate any of these particular variables (p. 146). A weakness of the ANOVA model is that the treatment effect is an average increase or decrease so the model does not describe the magnitude of responding (i.e., the extent of the impact). This cannot be rectified without adding other measured factors such as pre-test that take account of individual characteristics. Because of this it has less power to detect true differences than other methods.

However, ANOVA can also be used as a two-step process which employs both the pre- and post-test scores. First, the pre-test scores are analysed, and if not statistically significant, the groups are assumed to be equivalent (Cohen et al., 2007; Gay & Airasian, 2003). Then ANOVA was applied to the post-test scores. It was possible to use this form of analysis in the current research where pre-test scores were first analysed for variables of self-efficacy, PCK and pattern language to check for similarities between groups.

ANCOVA is an extension of ANOVA by including the pre-test in the form of a linear regression (Gersten et al., 2005). Using the pre-test covariate, it provides an adjustment for initial differences between groups (Verma & Mallick, 1999), which provides a benefit over ANOVA. To add further validity to the use of ANCOVA, a matching procedure is included where any given pre-test value is made into a predicted post-test value for treatment and control groups and the differences can be examined between them by use of a regression line. ANCOVA reduces the size of the error variance by including pre-test values directly in the model. It uses the pre-test as a predictor and assigns error only to the variance of the cross sectional scatter plots. ANCOVA has a smaller error variance than ANOVA. This difference in error variance depends on how highly the pre- and post-test are correlated within treatment groups. The higher the absolute values of the correlation, the greater the reduction in error variance (little reduction occurs until the correlation is above .3 or .4 though). ANCOVA models differential growth between groups only when the growth is perfectly mirrored by growth patterns that occur within the groups as the regression slope is used to correct for between groups differences.

Regardless of this benefit, ANCOVA still has biases that need to be addressed. When pre- and post-test are operationally identical measures, threats are lessened. However, it is rare in social sciences to have perfectly reliable measures as few behavioural states
can be measured without error. Use of ANCOVA with a single covariate is biased
towards under adjustment direction. A further source of bias occurs when considering
complex behaviours as change can be multivariate in nature. That is, behaviour may be
best described as a combination of separate independent characteristics which could
become disproportionately important at various times. ANCOVA can easily be
extended to include multiple pre-test measures or covariates so that adjustments can be
made simultaneously on several measures rather than just one. The treatment effect
estimate is still based on the difference between the regression line intercepts in the
two groups, but now the regression lines can be calculated using all pre-test measures
simultaneously.

This complex form of behaviour was illustrated in the current study with use of three
separate measures used at pre-test. These behaviours were not selected at random but
rather determined after careful analysis of the literature about factors that positively
influence inclusion. The inclusion of multiple covariates can increase the power of the
analysis and this benefit of increase in power was weighed up against each taking up a
degree of freedom from the error term.

It is also possible to increase the validity of the non-equivalent group design when
procedures such as blocking or matching are used. This is based on pre-test
performance after the groups have been formed. Individuals are grouped according to
agreed criterion and the pair assignment is then considered a factor in the data analysis.
The intuitive reason for blocking is clear, however a number of biases may still be
present.

When weighing up advantages and disadvantages of using ANOVA or ANCOVA
within the quasi-experimental design, it would appear that ANOVA was better suited
to the current research as it incorporated the pre-tests on multiple dependent variables
in linear regression, so the adjustments for initial differences on all measures were
included at the same time. As early analysis indicated the differences between groups
were minimal, the groups were considered as ‘equal’. By including the pre-test values
directly in the model it reduced the size of error variance. It is acknowledged that
behaviours such as ‘inclusive practices’ are complex constructs and difficult to define;
this bias is reduced by incorporating several dependent measures that relate as facets of
inclusive practice instead of relying on one.
3.4.1 Threats to validity and control over experimental designs.

Researchers are concerned with issues of reliability and validity regardless of the research design chosen. Researchers commonly distinguish between internal and external validity (Burns, 2000; Shaughnessy et al., 2009). Cook and Campbell (1979) assert a further two classifications are useful to consider: internal validity aligned with statistical conclusion validity, and external validity aligned with construct validity. Both will now be examined further.

3.4.1.1 Internal validity.

Internal validity is concerned with the notion that experimental treatments (independent variables) actually make a difference to dependent variables under investigation (Verma & Mallick, 1999). Once it is established that two variables co-vary, it needs to be determined if there is actually a causal relationship in existence as well as the direction of the causality. Internal validity is concerned with the threats or factors other than the independent variable that affect the dependent variable (Kervin, Vialle, Herrington, & Okely, 2006). These threats or rival explanations influence the outcome and are not part of the independent variable under investigation. The net bias these threats cause is dependent on whether they are similar or different in direction of bias, and on the magnitude of the bias they may cause individually (Cook & Campbell, 1979). It was appropriate to attempt to rule out these factors that posed ‘threats’ to the validity of the research results. However, maximising internal validity requires very rigid controls over participants and conditions. The more a study is narrowed, the less it is likely it is to be generalisable to realistic scenarios or applied settings. The researcher needs to strike a balance between control and realism, and between highly controlled and highly naturalistic environments (Gay & Airasian, 2003, p. 359). Utilising quasi-experimental design is one means of striking a balance (Lankshear & Knobel, 2004).

3.4.1.2 Statistical conclusion validity.

Cook and Campbell (1979) assert that statistical conclusion validity is closely associated with internal validity as it is also concerned with issues about covariation of the data. Statistical conclusion validity considers the distinction between bias and error where bias refers to factors that systematically affect the value of the means, and error refers to factors that affect the variability and thus decrease the chance of obtaining
statistically significant results. It is concerned with sources of random error and the appropriate use of statistics and statistical tests to reduce these threats. Researchers need to know if reasonable statements can be made about existing covariation and the strength of the covariance. The first question was concerned with the statistical power of the experiment given nominal sample sizes and degrees of variance. Particular sample sizes are required to detect an effect of desired magnitude. This can be computed according to formulae (Burns, 2000; Cook & Campbell, 1979). If an experiment has already been completed, the power analysis has a different function—that of determining the level of confidence the results can reasonably be relied upon. This is expressed in terms of confidence level and is set at 95% or 99%. Observed relationships that fall below the 5% confidence level are taken to be true rather than occurring by chance.

### 3.4.1.3 External validity

External validity, sometimes referred to as ecological validity, asks if the demonstrable effects may be generalised to other settings or populations (Mitchell & Jolley, 2007; Shaughnessy et al., 2009). For results to be useful, they may need to be generalisable beyond the confines of the study as well, and some assurance given that effects are replicable in other settings. Associated with external validity is construct validity, which is concerned with whether the measures used in the study actually measure factors that are supposed to be measured (Mitchell & Jolley, 2007). Similarities exist between construct and external validity in that they both indicate or codetermine a statistical interaction of the direction and size of a cause-effect relationship across populations.

In any quasi-experimental research there are factors that are not part of the planned intervention or design that may have an impact on the outcome. That is, there are uncontrolled variables effecting performance of the dependent variable (Gay & Airasian, 2003). Statistical conclusion threats, internal, external and construct validity threats will now be examined in detail in terms of the current research.

### 3.5 Threats to Internal Validity

Burns (2000), Cook and Campbell (1979), Fraenkel and Wallen (2006), and Shadish, Cook, and Campbell (2002) identify eight common threats to internal validity: history, maturation, statistical regression, testing, instrumentation, selection bias, and attrition,
as well as the interaction of these threats with each other. Shaughnessy et al. (2009) assert that threats of history, maturation, testing, instrumentation and regression are largely controlled for in non-equivalent control group designs. The potential for these threats to exist in this research and the means of lessening the effects are presented in Table 3.3 below and elaborated in the following sections.

Table 3.3

<table>
<thead>
<tr>
<th>Potential threats to internal validity</th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection Bias</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Maturation</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Attrition</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Additive and interaction effects (i.e., Selection – maturation interaction)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Instrumentation</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Statistical regression</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

3.5.1 Selection bias.

Differential selection of participants is an example of selection bias and a threat to validity when there are differences in characteristics between one group of participants and another before the commencement of treatment, especially if one group is to be used as a comparison or control (Gay & Airasian, 2003). These differences could be responsible for nuances in results at conclusion of the study. This is a major threat in quasi-experimental designs as it is impossible to randomly allocate participants to the experimental or control groups (Shaughnessy et al., 2009). Variables such as: the ability to do well on tests, age, anxiety levels, experiences with people who have a disability, learning style, cognitive ability, or being older and wiser all have potential impact on results and might occur differentially in one group compared to another.

In this research it was impossible to randomly assign participants to groups due to the nature of the educational settings from which they were drawn. The university classes were already formed across campus settings with 85 participants in one cohort and 24 in another. The cohort comprising 24 remained as one tutorial group in the applied experience cohort, and the cohort with 85 was further divided into four tutorial groups in the embedded design cohort. PSTs self-selected these tutorial groupings. The
researcher had no control over composition and there would be no guarantee that students in different tutorial classes or settings would have similar characteristics. However, the students did not self-select whether to be in the treatment condition or not as this was determined by the researcher. Two forms of analyses were conducted: two-way contingency table analysis and independent samples t-tests in order to reduce the threat of selection bias, pre-intervention differences were examined across conditions for each of the demographic variables.

When random selection is not an option, as in the current study, groups can be selected that are as similar as possible and a pre-test of dependent measures administered to check for initial equivalence. To reduce the threat of selection biases in study one of this research, pre-occasion data were collected on all dependent variables of interest as well as demographic information in order to check for nuances between the groups. The results at pre-test level for pedagogical content mastery, self-efficacy, and use of professional language indicated no significant differences and are reported in full in the results Chapter 4.

3.5.2 History.
Time between pre- and post-test observations or data gathering points may have an impact on results. Events other than the experimental treatments may occur during these times and mistakenly be attributed to the intervention or treatment alone (Gay & Airasian, 2003; Verma & Mallick, 1999).

Considering the naturalistic setting in which this research was conducted, one would expect that there would be many factors that could impact on PSTs’ self-efficacy, PCK, use and sophistication of professional language, and ability to implement strategies in the field in a manner with high fidelity. For example, there may have been instances where sickness occurred which could affect all variables under consideration if periods of time were missed from university study. It is foreseeable that these events might affect both groups equally and thus may be less likely to constitute a major threat. Wherever they are evident, they should be noted and the degree of their effect assessed if possible (Fraenkel & Wallen, 2006). The researcher needed to take these factors into consideration but also acknowledge that there are many outside influences that may affect PSTs within the timeframe of the data collection points that remain unknown and out of the realm of control.
3.5.3 Maturation.
Maturation is a threat to internal validity in terms of natural physical, emotional and
technological changes that occur over time as the participants grow older, wiser and gain
more experience. These changes could occur differently for participants in the different
groups. Usually it is considered a major threat in longer studies or studies with very young participants. During study one, the initial data for self-efficacy, PCK, and use of professional language was taken over a relatively short timeframe of six months so this threat was lessened.

Cook and Campbell (1979, p. 107) suggest utilising demographic information to determine if the non-equivalent groups are in fact maturing at different rates. If maturation threats are in operation, there should be an increased within group variance at the post-test as compared to the pre-test scores. Pre-test outcome scores were plotted against the maturation variables (i.e., age or prior experience) for experimental and control groups separately. Background data or longitudinal data was also used to reduce this threat. In the current research, aptitude was assumed as all participants gained university entrance to the teaching award and subsequent GPAs were also used as covariates and examined to determine starting differences. The threat of maturation to validity was also lessened as both groups were exposed to the same testing occasions.

3.5.4 Attrition/Mortality.
Attrition refers to the case where participants drop out of the research. It also becomes an issue when different groups drop out for different reasons and with different frequency (for example, participants may drop out if they become less interested than those who remain) (Kervin et al., 2006; Mitchell & Jolley, 2007; Verma & Mallick, 1999). This could eventuate from the experimental condition as well as the control group.

One way to assess the mortality of groups across study one and two was to obtain demographic information prior to the start of the study and then to compare the groups at the end. From this data, one can also identify the types of participants who did drop out and remove similar portions from the control group (Gay & Airasian, 2003). Some participants dropped out of the course or failed to complete some of the dependent
variable measures for reasons such as sickness on the day of administration, failure in multiple courses in the program or family issues. Even though the differences in sample sizes was not attenuated, this threat was managed in the current study by only incorporating full data sets for participants involved.

3.5.5 Additive and interactive effects to validity.
Not only can these threats to internal validity operate independently in a study, they may also operate simultaneously with each other and have an impact on outcomes of the treatments (Mitchell & Jolley, 2007; Shadish et al., 2002). Examples would include: selection-maturation interaction; selection-history interaction; and selection-instrumentation interaction (Gay & Airasian, 2003). Selection-maturation as an additive effect occurs when one group matures at a different rate to the other. Shaughnessy et al. (2009) suggest it is more likely to be a threat to internal validity when the treatment groups were self-selected. In this instance, the members of the treatment group were able to self-select tutorial groups but not whether they were to receive the treatment or not. It was determined entirely on campus location by the researcher.

As participants came from different settings and campuses, one might assume they also came with unique backgrounds. To minimise the interactive threat to validity, pre-intervention data was used to establish initial differences between groups and also differences at the conclusion of the study. Any results that indicated extreme values in the pre or post data occasions were excluded from the data set to minimise this threat. Thus, evidence strongly suggested the groups were close to equivalent at the time of pre-test occasions.

3.5.6 Testing.
Pre-tests can sensitise participants to the reason for the research and practice effects may produce higher scores on post-test measures than would have been the case had a pre-test not been taken (Mitchell & Jolley, 2007). This threat is sometimes referred to as pre-test sensitisation interaction and occurs when participants respond differently to a treatment simply because they have been pre-tested. Gay and Airasian (2003) suggest studies involving self-report measures such as attitude are particularly susceptible as participants may remember the questions and have had time to reflect on them in the interim timeframe and so answer differently than if new questions were
posed. They go on to suggest that studies conducted over a period of months or longer would probably have the effects of the pre-test worn off or be significantly diminished by the time the post-test is administered. Gay and Airasian (2003) suggest this is of greater threat when the time between pre- and post-testing is short.

Pre-tests were used extensively in study one of this research. In instances of the self-efficacy dependent variable, familiarity should be minimised as a threat and not have enhanced performance as the time interval was of six months between pre and post administration; and then a further 12 months during PST practicum placements. However, in the instance of PCK, the pre-test was followed by successive quizzes held at intervals between three to four weeks’ duration and on similar content to the pre-test as each of the content areas was covered in class. This potentially lent itself to sensitisation and, in fact, compounding, and was conducted intentionally as a part of the mastery learning paradigm. Newly acquired knowledge was assessed at each occasion and compared across conditions. The testing for pattern language occurred where the PSTs were asked to write a reflection on their learning and this was collected for later analysis. In study two, the testing of pattern language occurred when PSTs were writing a reflection about the design and delivery of their lesson.

3.5.7 Instrumentation.

Lack of consistency from unreliable tests or instruments can introduce serious errors into results. If two different tests are used in pre- and post-tests, instrumentation may become a threat (Gay & Airasian, 2003; Mitchell & Jolley, 2007). In this study, all instrumentation used as pre and post occasion means of obtaining data were the same in order to minimise this threat. The measures have also been used in previous work where reliability and validity were established (Bain, Lancaster, Parkes, & Zundans, 2009; Bain, Lancaster, & Zundans, 2009; Lancaster & Bain, 2007).

One standardised measure was used in study one where the performance was measured against a normed group. The norming sample of the SEIPD was an Australian group of undergraduate teachers just as in this research sample. If standardised instrumentation is problematic in delivery of results, it may not be clear if the intervals used in standardised scales are equal, where change may be easier to detect at some points than others. The SEIPD used in study one and two has reported statistical reliability; and the group mean differences at pre and post occasions were checked for skewness.
The other criterion referenced tests that were specifically created for study one did not have these reliability coefficients to report. In instances of criterion referenced tests (CRTs), PST performance was based on curriculum specific tasks developed for the study, and any effects of intervention were measured by systematic progress monitoring in the form of quizzes (Howell & Nolet, 2000, p. 143). As opposed to the norm referenced test, the criterion referenced test was designed to measure a narrow band of curriculum. To determine if PSTs had sufficient knowledge of explicit teaching and cooperative learning pedagogies, the CRTs criteria for passing was set at 80%. CRTs lend themselves to repeated use and can determine PST progress towards a particular goal when criteria are set in this way. All of these standards applied to the development of CRTs used in study one, and so the criterion referenced instruments used for dependent measures were not deemed to be highly problematic in this research.

3.5.8 Statistical regression.

This source of error refers to regression towards the mean which is caused by participants scoring high mean levels on a pre-test and thus leaving little room to improve on the post data collection occasions (Cook & Campbell, 1979; Gay & Airiasan, 2003; Shadish et al., 2002). The magnitude of regression depends on the test-retest reliability found. An example of this statistical source of error may be referred to as the ‘ceiling effect’ (Cook & Campbell, 1979; Gay & Airiasian, 2003; Shadish et al., 2002) where very high scores are obtained in the pre-test data collection occasion and thus not leaving sufficient room to demonstrate actual improvement by post-test. ‘Floor effects’ may also occur when scores from one group cluster around the lower end of the scale. These instances are considered in the analysis of the data especially calling upon the pre-occasion data to determine starting points on the various dependent measures being investigated. Whenever participants are selected on the basis of their extremely high or extremely low scores, statistical regression is a viable threat. In study one this had the potential to impact on measures of self-efficacy, PCK, and use of pattern language. When it was noted that participants were scoring extremely high or low scores (outliers), they were rejected from the analysis to prevent this from skewing results. Thus the threat to validity due to statistical regression was lessened.
3.6 Threats to Statistical Conclusion Validity

These threats are concerned with whether the independent and dependent variables are actually related (Mitchell & Jolley, 2007; Shadish et al., 2002). Various types of statistical conclusion validity include: statistical power, violated assumptions of statistical tests, error rate issues, reliability of measures, reliability of instrument implementation, variance in the settings, and heterogeneity of participants. Burns (2000) suggests we need to consider three closely related issues: that of power, effect size and types of errors when interpreting results. Each is now elaborated in terms of the current study.

Table 3.4
Potential Threats to Statistical Conclusion Validity

<table>
<thead>
<tr>
<th></th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low statistical power</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Violated assumptions of statistical tests</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Error rate problem</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Reliability of measures</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Reliability of treatment implementation</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Extraneous variance in the experimentation setting</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity of respondents</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

3.6.1 Low statistical power.

The power of a statistical test is its ability to detect differences to see if the research hypothesis is true, and the probability of making a correct rejection of the null hypothesis when it is false. Factors that positively influence statistical power include: effect size, differences between the population means, small population standard deviations, and large sample sizes, lenient significance levels set, and one-tailed versus two-tailed tests. The bigger the difference obtained on mean scores between two groups, the more power in the study; it is also true that the smaller the variance within the two groups, the larger the effect size. The greater the effect size, the greater the power. Measures of effect size are used to determine the amount of impact of the independent variable rather than simply stating there was an effect. Means from experimental and control groups in study one were determined, and thus the effect size calculated. The extent to which the two populations did not overlap is the effect size,
as it is the extent the manipulation has had on separating the two groups. Effect sizes range from small (0.2), medium (0.5), and large (0.8) (Burns, 2000).

Statistical power is important when planning a study as it helps determine the number of participants needed. This is of particular importance in the current study due to the differences in sample sizes involved. The sample size can be calculated according to formulae. Once a study has been completed, power analysis takes on a different function. The known variances and sample sizes are then used to determine if the effect that could reasonably be detected (i.e., with 95% confidence level). The degree of random error acceptable is usually at the 0.05 level (i.e., 5%). The likelihood of making an incorrect no-difference conclusion (type II error) increases when sample sizes are small and alpha is set at a low level. Sample size in the current study was 109 across the two settings and significance levels were set at 0.05 level to reduce this threat to validity. In order to achieve an effect size of .2 (small effect) with a power of .5 in the current research, the requirement of a sample size of 100 participants was met (Burns, 2000, pp. 173-174).

3.6.2 Violated assumptions of statistical tests.
Most tests require that certain assumptions be met if the results are to be meaningful. Many of these are listed below, for example, the sample size required, the variances within the groups being compared, the used of one or two tailed tests, the significance level selected, the number or participants required in the sample, and the desired power. If they are not met it may not be valid to draw conclusions from the results. For example: the examination of scatter plots and comparing pre-test means would determine if the groups being compared were equivalent. Having said this, not all assumptions are of equal importance. The analysis of variance (ANOVA) technique is robust to violations of normality (Cook & Campbell, 1979). More detailed analysis of particular test assumptions follows with descriptions of the specific tools being used.

3.6.3 Error rate problem.
Error rate problems relate to situations where the wrong conclusions may be reached even when everything was carried out correctly. This threat to validity occurs where conclusions are made that covariation exists when it actually does not (Type I error). Type I errors relate to an incorrect conclusion that a difference exists and may arise when the significance cut off is set too high (say 0.2). This level would allow the null
hypothesis to be rejected too easily. To reduce the chance of making a Type I error, the significance level may be set lower so there is less chance of the null hypothesis being rejected so easily.

Type I errors may also occur when there are multiple comparisons to the mean differences possible. Error rates per comparison needed to be considered. New $t$ values can be computed by dividing it by the number of total comparisons to be made to give an adjusted proportion for ($p$). Another method of dealing with error rate problems is to use a more conservative multiple comparison test such as Scheffe. Scheffe may be employed to check for covariance of variables in study one with the implementation variable of study two.

Type II errors are related to an incorrect no-difference conclusion, where significance has been set too low and results do not come out extremely enough to reject the null hypothesis when it perhaps could have been. The compromise here is to incorporate more lenient significance levels than the standard 5% or 1% levels which was the approach taken in this research.

3.6.4 Reliability of measures.
This is conceptualised as reliability or stability of the measures used. The standardised measure used (SEIPD) had high test-retest coefficients of 0.8 over a four-week period, which met accepted standards. If this stability is not considered, unreliability could inflate the standard errors of estimates and thus accuracy in inferring differences between the means of the groups. The other criterion-referenced dependent measures used in the study were checked for reliability with the assistance of additional scorers. The three additional scorers were: the second tutor working with the researcher for the embedded design cohort, the tutor for the applied experience cohort, and an assistant scorer who had no knowledge of the research questions. At each data collection point for dependent measures, both pre and post occasions, a random selection of the participants’ data were scored and interrater agreement achieved between the researcher and research assistant at above 90%. During study two, a random selection of the participant data was also scored by both researcher and research assistant.
3.6.5 Reliability of treatment implementation.

The variations in the way the treatment condition (the independent variable) is implemented can inflate error variance and may decrease the chance of obtaining true differences (Burns, 2000; Fraenkel & Wallen, 2006). Within the context of this study, internal validity may be threatened by groups being exposed to different lecturers. Two tutorial groups from the embedded design cohort were conducted by the researcher; two further tutorials were conducted by a colleague. The lectures and tutorials for the applied experience cohort were conducted by another colleague, in all a total of three instructors. The implication here is that consideration needed to be made of the way the course may be taught in different ways by different personnel. This was considered a threat in this research and was reduced by trying to standardise the treatment across tutorial groups for the embedded design cohort. The applied experience cohort was not utilising embedded design in delivery (rather an applied experience design) of the course so close attention was required to report exactly the way the course was delivered on each campus. The following is an account of procedures put into place to document treatment implementation fidelity across campuses. Data collection for the course implementation fidelity may be found in Appendix 1.

The course implementation fidelity table was constructed by the researcher. Each week all PowerPoints used in lecture delivery and activities planned to be covered in tutorials were sent to the corresponding instructors for the course. Towards the end of the week, the researcher emailed a table document that summarised these activities and content where a scale was completed ranging from: Strong agreement = (yes I did all of this during the week); moderate = (I did some of this during the week); and weak = (I did not do this during the week).

Where there was moderate or weak agreement, further details were sought and recorded by the researcher (These are noted in Appendix 1). This was a measure of how individual lecturers felt they covered the intended content at each site. These findings are reported in the results chapter (See section 4.1.1).

3.6.6 Extraneous variance in the experimentation setting.

This involved the interference of extraneous sources apart from the treatment variable in the settings concerned. In naturalistic settings it is impossible to control sources of variation. It was not possible to control for extraneous variance due to the nature of
the settings where the research occurred. For example, tutorial groups had different instructors responsible for tutorials across cohorts. While locations cannot be standardised, implementation procedures were carefully documented and data collection and analysis standardised. Careful monitoring of settings in terms of course delivery is documented in Appendix 1: Course implementation data collection tool across cohort delivery.

3.6.7 Heterogeneity of respondents.
Some of the participants may have responded to the treatments differently than others. This threat is sometimes referred to as ‘subject characteristics’ (Fraenkel & Wallen, 2006), where characteristics such as critical thinking ability pose a difficulty unless controlled by the use of matching procedures, etc. These authors also suggest gender is an issue and thus demographic information was collected in the current research at the pre-test occasion of study one in order to determine the effect. This threat was reduced by considering within group pre and post results as opposed to differences between respondents. Cook and Campbell (1979) suggest that the higher the correlation between the pre and post scores, the less likely this type of error would have an impact.

3.7 Threats to Construct Validity
Construct validity is concerned with the outcome construct as well as the treatment construct (Cook & Campbell, 1979; Shadish et al., 2002). It is concerned with the question: Do the measures used in the study actually measure factors that are supposed to be measured? It infers that there should be checks and balances to ensure the independent variable actually alters what it is supposed to alter. Does the independent variable have an impact on related constructs?

The proposed dependent variables should also represent the factors they are meant to measure and some form of inter-item correlation might suggest this. The dependent variables should not be dominated by irrelevant factors. Unfortunately it is difficult to have concise definitions of complex constructs either in the form of independent variable treatments or dependent measures assumed to be related to successful intervention. Construct validity is seen as “the fit between operations and referent constructs” (Cook, 2002; Cook & Campbell, 1979, p. 63) and thus required a rigorous
definition of each. All the dimensions of the construct need to be considered and irrelevant dimensions of the construct omitted.

Table 3.5
Potential Threats to Construct Validity

<table>
<thead>
<tr>
<th></th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate operational explication of constructs</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>The Hawthorne Effect</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Experimenter effects</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Confounding constructs and levels of constructs</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

3.7.1 Inadequate operational explication of constructs: Independent variables.
The independent variable is used as the experimental treatment of one group compared to another where the treatment was withheld or a different one applied. In study one of this research, the independent variable remained the same and is known as the application of the embedded design principle. Gay and Airasian (2003) suggest that specificity of the variables is a potential threat regardless of the experimental design used. Specific participants, operational definition of independent and dependent variables, at a specific time, under specific circumstances all need to be described adequately so the study may be replicated. Readers also need to be able to assess if the findings are applicable in their situation.

Conceptual analysis of essential features should be a starting point here. “A precise explication of constructs is vital … as it permits tailoring the manipulations” in a more precise manner (Cook & Campbell, 1979, p. 65). This sort of validity was sought by operationalisation of the independent variable constructs under consideration. To counter this threat in the current research, considerable time was devoted to defining the principle of embedded design and also applied experience designs used by the cohorts. The researcher was in contact weekly with colleagues to determine exactly what was covered and the manner in which it was covered.

3.7.2 Inadequate operation of the dependent variables.
Dependent measures are the variables that will demonstrate an effect after implementation of the experimental treatment. As per discussion above for
operationalisation of independent variables, validity is also sought by operation of the dependent variables under consideration. Unless dependent variables are adequately described, future replications of the study are impossible. Gay and Airasian (2003) caution that when there are a number of dependent variables in use, questions about comparability need to be raised. High levels of reliability scores taken by the researcher, colleagues, and a volunteer researcher indicated the operationalisation of the dependant variables was adequate. Reliability testing was conducted for study one dependent measures at pre and post-test occasions. These included dependent measures of self-efficacy, PCK and pattern language frequency as well as sophistication.

3.7.3 The Hawthorne effect.
This occurs when participants realise their role in research and the mere participation in experiments may bias the results found. Simply being in the study may cause different behaviour to the usual. In order to counteract this in medical studies, double-blind designs may be employed; however this sort of procedure was not possible in naturalistic settings of the current research. The participants in this study appeared to be treated the same as they were all enrolled in the same course, with the same objectives set as a requirement to be met. The threat is minimised when dependent measures are taken on a regular basis and treated as a form of constant feedback for PSTs as part of their course design. For example, in study one, pre-test occasion of PCK was followed up with quizzes where the most recent topics were tested for mastery and feedback given to PSTs in order to reach the designated mastery. The Hawthorne effect was also minimised as the treatments occurred in naturalistic settings.

3.7.4 Experimenter effects.
Researchers themselves may present potential threats to external validity in terms of unintentionally effecting the study procedures, the behaviour of the participants, or the assessment of their performance. Experimenter effects may be passive or active (Gay & Airasian, 2003). Passive elements included things such as age, gender, race, and anxiety level. Active effects may come about due to researcher expectations of the study results.
At the time the study occurred, there were three instructors involved with the inclusive education course. These included the researcher, who had 10 years of university teaching in the area of inclusion; 11 years in the field of expertise and 7 years of mainstream teaching at the primary school level. A second staff member who was a tutor for the embedded design cohort who had 2 years of university teaching in the area of inclusion at the current university; 15 years in the field of expertise at the secondary school level, and 10 years in mainstream classrooms. The final staff member working with the applied experience cohort had 5 years of university teaching in the area of inclusion at the current university; 2 years at TAFE, and 13 years in the field of expertise at the primary school level. The highest level of academic qualification achieved for each of the instructors was Masters of Special Education.

These sorts of biases may be minimised by actions that prevent evaluating results differently. For example, scoring dependent variables ‘blind’ was practiced for all variables under consideration. The effects can be minimised by utilising experimenters who have no expectations or false expectations, or by analysing the data separately to those directly involved. Unfortunately this was not possible in the current research where the principal researcher was directly involved in all stages of the research. Where other teaching personnel were involved, they were also aware of the purposes of the research. The research volunteer was not aware, however, and this assisted with reliability coding of data. Obtaining reliability on all dependent measures as well as rigorous implementation evaluations in study one and two helped to minimise this threat.

3.7.5 Confounding constructs and level of constructs.

There may be several discrete levels of independent variable that are continuous, thus covariance at each level was measured against the dependent variable to determine effect. In this research, the independent variable was described as consisting of two levels; and the embedded design principle consisting of a further four defining characteristics. These were described in detail in order to operationalise them. However it must be noted that they were not treated as separate, as the intervention involves the interaction of each of these levels being used on a cyclic basis, so to separate them out for individual analysis defeats the purpose of the embedded design feature.
To account for construct validity, this research carefully defined the constructs of interest, isolated cognate characteristics from which the particular construct of interest was differentiated, and measures were determined that could be used to index the construct of interest. Multiple measures were desirable wherever possible as no single measure is sufficient or adequate. All measures used to determine impact of the independent variable were sourced from the research literature available on inclusive practices.

3.8 Study One Procedures

Participants, conditions and details of the teaching cycles will be elaborated first for the *embedded design* and then the *applied experience* cohorts.

3.8.1 Participants.

Table 3.6 provides a summary of the demographic information collected about the participants in study one. There were no statistically significant differences between the cohorts for each demographic variable. In study one, there were slightly younger PSTs in the *embedded design* condition (N=72) 85% of participants were under 25 years old; in the applied *experience condition* (N=16) 68% fell into this age bracket. Their academic abilities were very similar, with GPAs to date being 4.65 for the *embedded design* and 4.66 for the *applied experience*. The average total points accumulated in the program to date were similarly close, with *embedded design* being 123.61 and *applied experience* being 125.45.

Ninety-five percent (N=81) of PSTs in *embedded design* were enrolled in their first degree at university compared to 100% (N=24) of PSTs from the *applied experience* cohort. Participants were asked about their level of experience with people who have disabilities. With experiences ranging from teacher aide work or practicum experience to having a family member or friend with a disability: more than 76.5% (N=65) in the *embedded design* had some level of experience as did 75% (N=18) from the applied *experience* cohort. Many similarities are evident from this demographic information provided about the two groups.

A total of 109 PSTs participated in study one, 85 from the *embedded design* cohort and 24 from the *applied experience* cohort. Of the total, 28 were male and 81 were female. The participants comprised the total population of students enrolled in the second year
of the Bachelor’s Degree in Primary (Elementary) Education at an Australian regional university. The students were enrolled at two campuses (described here as the *embedded design* cohort and the *applied experience* cohort), located in regional Australian cities 200 kilometres apart.

Two-way contingency table analyses were conducted to detect pre-intervention differences between conditions on demographic variables including: age; previous experiences with individuals who have a disability; and degree of teacher training completed ($\alpha = .05$). Prior to running the analyses, the following assumptions were examined: (i) independence between variables; and (ii) homogeneity of proportions. Both assumptions were met for each of the demographic variables with the exception of assumption two for degree of study.

Independent-samples $t$-tests were conducted to detect pre-intervention differences across intervention conditions in demographic variables including: (i) grade point average (GPA); and (ii) total points accrued in the teacher education program. In addition, $t$-tests were also conducted to detect differences in pre-test performance on all dependent measures including: SEIPD; pedagogical content knowledge; and pattern language frequency.

### 3.8.1.1 Instructors

At the time the study occurred, there were three instructors involved with the inclusive education course. These included the researcher, who had 10 years of university teaching in the area of inclusion; 11 years in the field of expertise and 7 years of mainstream teaching at the primary school level. A second staff member who was a tutor for the *embedded design* cohort who had 2 years of university teaching in the area of inclusion at the current university; 15 years in the field of expertise at the secondary school level, and 10 years in mainstream. The final staff member working with the *applied experience* cohort had 5 years of university teaching in the area of inclusion at the current university; 2 years at TAFE, and 13 years in the field of expertise at the primary school level. The highest level of academic qualification achieved for each of the instructors was Masters of Special Education.

The researcher was responsible for lecture design and delivery for the *embedded design* cohort, as well as two of the tutorials for the *embedded design* cohort. The
second instructor was responsible for the remaining two tutorials of the *embedded design* cohort. The third instructor was responsible for the lecture delivery as well as the tutorial for the *applied experience* cohort.

Table 3.6
Demographic Information of Participants in Study One

<table>
<thead>
<tr>
<th></th>
<th>ED</th>
<th>AE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 20-25</td>
<td>72 (85)</td>
<td>16 (66.66)</td>
</tr>
<tr>
<td>Age &gt; 26</td>
<td>13 (15)</td>
<td>8 (33.33)</td>
</tr>
<tr>
<td>Cohort</td>
<td>85 (77.98)</td>
<td>24 (22.02)</td>
</tr>
<tr>
<td>GPA</td>
<td>4.65</td>
<td>4.66</td>
</tr>
<tr>
<td>Average total points accrued in program</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>First degree of study</td>
<td>81 (95.29)</td>
<td>24 (100)</td>
</tr>
<tr>
<td>Some experience</td>
<td>44 (51.76)</td>
<td>12 (50.00)</td>
</tr>
<tr>
<td>Little experience</td>
<td>21 (24.70)</td>
<td>6 (25.00)</td>
</tr>
<tr>
<td>No experience</td>
<td>20 (23.52)</td>
<td>6 (25.00)</td>
</tr>
</tbody>
</table>

The technique of opportunity (or convenience) sampling was invoked as the researcher was teaching the *embedded design* cohort of second year PSTs at the time (Burns, 2000). Participants were already in tutorial groups across settings so they formed conveniently accessible groups. As this method of sampling potentially involved considerable possibility for error, it was considered essential that additional information be collected from the participants to establish similarity and differences between the groups.

3.8.2 Conditions.

In study one, the teaching sessions comprised a 13-week course with a total contact time of 39 hours, held in the lecture theatres and tutorial rooms across two university campuses. The *applied experience* condition had 11 hours of additional time visiting in-situ settings, bringing the total contact time for students in the *applied experience* condition to 50 hours. These experimental conditions for the independent variables are referred to as the *embedded design* condition and the *applied experience* condition. Lectures were of one hour and included all students, while tutorial/workshop sessions
were of two hours and included approximately 20 PSTs in each class. The differing designs and teaching cycles are described below for each cohort.

3.8.3 Embedded design and levels of the independent variable.
The type of ‘course design’ the students participated in distinguished the two levels of the independent variable. They were: embedded design and applied experience design. Embedded design is a critical component of self-organizing systems theory that has been applied to educational contexts (Bain, 2007). Here it applied to course development for PSTs in an undergraduate elementary teaching degree. Applied experience utilises additional time in-situ with school students to work with them and improve their literacy skills. Both embedded design and applied experience are elaborated below for clarification.

3.8.3.1 Self-organizing systems theory.
The self-organizing approach involves applying principles derived from prior work on a theory of self-organizing schools (Bain, 2007) to teacher education course design. Self-organizing systems theory involves the application of six principles that enable change from the ground up which include: simple rules, embedded design, similarity at scale, emergent feedback, dispersed control and use of a schema to guide action and accommodate change (Bain, 2007). These principles used for self-organizing schools were described in detail in the review of literature in Chapter 2.

Any successful complex systems exhibit self-repeating patterns within their organisational structure (Waldrop, 1993). The focus of this study involved investigation of the effect of one of these principles, that of embedded design as it was applied to a course within a PSTs’ primary education program. Embedded design involved creating these self-repeating patterns in a system by expressing the essential features of the system at many levels in its design (Bain, Lancaster & Zundans, 2009). Embedded design allowed for the inclusion of common features to courses such as: self-questioning, peer mediation, authentic assessment, advance organisation, and concept mapping into a research driven framework (Bain, 2007; Bain, Lancaster, Zundans & Parkes, 2009).
3.8.3.2 Embedded design.

In this research, *embedded design* represented a means to create a schema to guide action. As Bain (2007, p. 107) explained: “It provides a forward mapping process that translates the assessment needs and simple rules into a … design”. Applying *embedded design* in this study meant employing the simple rules that articulated research-based teaching and learning strategies. It assumed that key features and content were embedded in all parts of the design. The *embedded design* principles were used as the treatment implemented for one *embedded design* cohort and were enacted via four different components: knowledge and awareness (quizzes); active experience (lesson plan designs); continuous feedback (from peers); and personal impact (grades achieved as a result of exposure to the first three levels described). Following is a detailed description of each component as applied in the present studies.

3.8.3.3 Component 1: Knowledge and awareness (Level 1 of *embedded design*).

Specialist PCK is identified in the literature as being essential for students who have special needs. This constitutes the first two steps identified by Bain (2007, p. 106) of *vision building* and *articulating simple rules* based on needs assessment of a context. The need for these strategies to be deeply understood is identified in the literature as they are rarely implemented with integrity or involving the necessary characteristics for success. Previous work substantiates the premise that application of *embedded design* covaries with high levels of mastery in the knowledge of pedagogy (Bain, Lancaster, Zundans & Parkes, 2009). In study one it translated as follows: all PST students were required to complete pre-reading on the following research-based strategies in preparation for lectures: collaborative problem solving, explicit teaching, cognitive strategy training, and cooperative and peer assisted learning. Lectures were then used to develop and apply the concepts and ideas described in the readings. PSTs in the *embedded design* cohort attended lectures over the 13-week period. The readings and lectures were threaded together by a set of specific objectives provided to PSTs during the week prior to the introduction of a new topic. The objectives explained the key understandings for each topic as well as how related information would be provided either in reading, by lecture, or both. PSTs were accountable for developing responses to each of the objectives for each week. Quiz questions were based upon the objectives. (See an example of sequence in Appendix 3.)
3.8.3.4 Component 2: Active experience for embedded design.

At this level of course design, workshops were used to translate knowledge and skills gained in tutorials into a series of practical experiences, and thus to deepen the learning experience. PSTs participated in five two-hour skill-building workshops. Workshops were conducted in the areas of collaboration, explicit teaching, cognitive strategy training, and cooperative learning. PSTs were taught how to build lesson designs using each of the approaches and then to differentiate those designs for an inclusive classroom. In each case, the teaching approach that constituted the topic of the workshop was employed to teach the workshop. For example, PSTs learned about cooperative learning by using cooperative learning (i.e., Jigsaw II-Slavin et al., 1996) as the medium of instruction in the workshop. The same approach was applied to the design and implementation of workshops on explicit teaching, and cognitive strategy training. (See example in Appendix 3).

3.8.3.5 Component 3: Continuous feedback for embedded design.

Continued feedback is essential to assist teachers to build a framework for instantiation of the research-based approach in practice (in the case noted above, into an active experience). The capacity to progress depends on feedback (Bain, 2007, p. 112). Feedback allows the schema to evolve as the capacity of the participants grows. The embedded design principle calls for the embedding of the key elements in all others (Bain, 2007) and providing feedback along the way for participants to ensure mastery. This was accomplished by using the collaborative problem solving process in all workshops as a medium for learning about other approaches. In the first workshop meeting (Week 2), PSTs were randomly placed in collaborative practice communities for the duration of the course and they learned a collaborative problem solving process together (Friend & Cook, 2013), practicing it first with simple problems like ‘naming’ their community. The application progressed to more sophisticated instructional problem solving related to the lesson designs.

PSTs convened their communities as a part of the teaching cycle for each inclusive approach in order to share their lesson designs. They shared copies of their designs with peers. After reading the design, the group used the collaborative process to provide feedback on each lesson. This process embedded collaboration in the learning about all other practices and called upon PSTs to make active use of their knowledge of the pattern language of explicit teaching, cognitive strategy training, and
cooperative learning by deploying their knowledge of those practices in the feedback exchange.

3.8.3.6 Component 4: Personal impact for embedded design.
At the personal impact level, embedded design has a direct, ‘non-simulated’ effect on the PSTs’ engagement with the course. PSTs use the inclusive practices in ways that have consequences for their assessment performance in the course. This involved using the inclusive practices taught from week to week as part of the PST preparation for their assessment tasks. PSTs met in cooperative groups in those class sessions in which quizzes were scheduled. For twenty minutes prior to the administration of the quizzes, the PSTs used the collaborative processes to prepare. As such, their capacity to employ the research-based characteristics of the inclusive approaches influenced the quality of their preparation and ultimately their quiz grade (Bain, Lancaster, Zundans, & Parkes, 2009). The embedding was intended to result in a more visceral or direct level of impact where students could experience, authentically, the effect of the approaches on their own learning and performance. Further, the PST lesson designs described in the previous section were also graded as an assessment requirement. The quality of the collaborative feedback each PST received from their peer group community influenced the quality of their revisions and in turn influenced the grade they received. The personal impact level of embedding occurred on three occasions for quiz preparation and on four occasions for lesson feedback in the course schedule.

Previous work substantiates the claim that application of embedded design at the four levels described above covaries with increases in student mastery of PCK and use of professional pattern language to describe these practices. Self-efficacy in working with students who have special needs also improved (Bain, Lancaster, & Zundans, 2009; Bain, Lancaster, Zundans, & Parkes, 2009; Lancaster & Bain, 2007).

3.9 Threats to External Validity
Cook and Campbell (1979) suggested that external validity is a two-step process. First, the target population, settings and times need to be defined, and then samples from this predefined population drawn. External validity refers to the extent to which the presumed causal relationship of a study can be generalised across groups of participants, settings and times (Shaughnessy et al., 2009). As true representativeness of populations is rarely possible in social science research, a convenience sample was
used. In relation to this study, one may be able to generalise results to other universities and teaching inclusive education strategies to PSTs. Further generalisation would require research that spans national and international settings. Cook and Campbell (1979) recommend that generalising to well-explicated populations should be clearly distinguished from generalisation across populations.

Burns (2000) identified six threats to external validity: failure to describe the variables explicitly, lack of representativeness of available populations, the Hawthorne effect, inadequate operationalisation of the dependent variables, sensitisation to the experimental conditions, and interaction effects of extraneous effects and experimental conditions. Cook and Campbell (1979) elaborate upon the final factor as: interaction of selection and treatment, interaction of setting and treatment, and interaction of history and treatment as being tests of statistical interactions. The initial threats identified by Burns were covered above under the heading of Construct Validity. The remaining threats are dealt with here. The following table summarises the extent each impacted on the current study.

<table>
<thead>
<tr>
<th>Threats to External Validity</th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of representativeness of available populations / selection treatment interaction</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Interaction of setting and treatment</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Interaction of history and treatment</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

3.9.1 Lack of representativeness of available populations/selection-treatment interaction.

This threat is similar to the differential selection of participants as noted earlier (Gay & Airasian, 2003) where participants are not randomly selected for treatments. The participants in this study are representative of an ‘available’ or ‘target’ population; they may not be representative beyond this setting. This threat is related to whether the relationships determined may be generalised to other tertiary settings and PST students
Apart from the assumption that the results may generalise to other PSTs, there may be characteristics of these PSTs or the staff who taught them in this study that prevent this generalisation from occurring. Sufficient detail is given about the participants, staff, and definitions of independent and dependent variables, that the study may be replicated.

The settings involved in this study were located in two university campuses in regional Australia. The treatment may have had specific results for these particular settings thus it may be difficult to generalise findings to future higher educational settings. Some authors recommend a solution to this threat by way of varying the settings and the treatment that is applied, and examining the causal relationships in each (Johnson & Christensen, 2008; Mitchell & Jolley, 2007; Shaughnessy et al., 2009). Preliminary studies may also be of assistance where they can be replicated (Bain, Lancaster, & Zundans, 2009; Bain, Lancaster, Zundans, & Parkes, 2009; Lancaster & Bain, 2007).

3.9.3 Interaction of history and treatment.
Threats in this category are posed by events that occur while the study is taking place but that are not related to the study. Events extraneous to the study may alter the results in any case. A similar threat is identified as the interaction of time of measurement and treatment effect where the administration of the post-test too soon after the treatment may not have the same effect as if a delayed post-test was administered. Thus, the only way to assess generalisability is to measure the dependent variable at various times following treatment. Cook and Campbell (1979) suggest solutions to short term historical event may be found by replicating the experiment at different times or by conducting a literature review to determine if prior evidence exists about the causal relationship. Both of these avenues were taken into account in the current research. Once all sources of threat were examined and minimised in the research as much as is possible, research questions were then developed and hypothesised.

3.9.4 Teaching cycle for the embedded design cohort.
During study one, the four levels of embedding were implemented sequentially for each topic and framed the week-to-week teaching cycle for the course. The cycle included pre-reading, lecture, skill building workshop, lesson draft development,
collaborative feedback, lesson submission and quiz. Each component of embedding focused on reinforcing the learning experience acquired at other component levels. For example, the approaches to cooperation (Slavin et al., 1996) used in quiz preparation were the same approaches that PSTs read about and that were described in lectures. The collaborative process used in class to review lesson designs was the same process introduced in the active experience workshop. In this way, each level of embedding was designed to have a self-reinforcing effect on the others as PSTs’ learning experience at one level was reinforced at another (Bain, 2007).

The table in Appendix 3 outlines ways the principle of embedded design was used in a cyclic fashion for each topic introduced. A lecture was given as an introduction followed by a workshop where the new strategies were modelled and the PSTs began to design their own lessons with guided practice. Feedback was incorporated the following week when PSTs presented drafts to their collaborative groups and independent practice was the culmination when the final lesson was submitted for grading (see Appendix 2). Within the cycle, feedback was also obtained on the mastery of PCK when the PSTs completed a quiz on the same strategy they were using to design a lesson for submission. Feedback from the quizzes was always returned the following week and prior to lesson design submission (see Appendix 3).

3.9.5 Conditions for the applied experience cohort.

Twenty-four participants in the applied experience cohort were taught by a colleague of the researcher; the two had worked together for many years in a cross-campus mode. The same time periods were applied to course delivery with lectures of one hour and tutorials of two hours duration. The same course objectives were covered by both cohorts as the course was part of the same degree, simply delivered in different settings. The first assessment item was identical, which involved the quizzes as well as pre and post knowledge tests. This assessment item constituted 30% of the grade for the course. Differences occurred for assignment two where the PSTs in the applied experience cohort were required to work with school students. Instead of completing the workshops completed by PSTs in the embedded design cohort in preparation for handing in their lesson plans, these applied experience students had an assignment based on their in-situ experiences with school-aged students. Assignment two constituted 60% of the grade for the course. The remaining 10% of grades were attributed to contributions and attendance at tutorials across both cohorts.
While different in their form or delivery, the approaches of each cohort met the requirements of the common university-approved profile for the course which existed to ensure that key knowledge and understandings would be addressed in all iterations. The applied experience cohort did not have the continuous application and feedback dimension or the personal impact dimension. The first two components (knowledge and active engagement) are not unlike what we would expect to see in any existing well organised university course.

3.9.5.1 Applied experience as an independent variable.

Level I: Knowledge and awareness.

Exactly as per the embedded design cohort, all PSTs in the applied experience cohort were required to complete pre-reading on collaboration, explicit teaching, cognitive strategy training, and cooperative learning, in preparation for lectures. Lectures were the same across conditions and were used to develop and apply the concepts and ideas described in the readings. PSTs attended lectures over the 13-week period. The reading and lectures were threaded together by a set of specific objectives provided to PSTs the week prior to the introduction of a new topic. The objectives explained the key understandings for each topic as well as how related information would be provided either in reading, by lecture, or both. PSTs were accountable for developing responses to each of the objectives for each week and quiz questions were then based upon these objectives. This level of embedding was implemented sequentially for each topic and framed the week-to-week teaching cycle for the course. The cycle included pre-reading, lecture, and quiz. It covered topics of: explicit teaching; cognitive strategies; cooperative learning; and differentiation of these strategies. PSTs from the embedded design cohort used collaborative problem solving and learning in preparation sessions to prepare for the quizzes they would take as part of their assessment. Those from the applied experience cohort were also allowed the same amount of preparation time prior to each of the quizzes. Unlike the structured group activities employed for embedded design, those in the applied experience cohort were allowed to choose how they prepared for the quiz. Consequently, some from the applied experience cohort chose to study independently, some chose to study in pairs, and some chose to study in groups. Those that did meet in groups were not necessarily using principles nor structured steps of collaborative practice. Quiz preparation occurred on three occasions.
3.9.5.2 In-situ experiences.

It is from here that the implementation substantially varied between cohorts. A major difference was seen where tutorials/workshops for PSTs in the embedded design cohort were used to translate knowledge and awareness into skill in a series of six practical experiences with the aim of building lesson designs. PSTs in the applied experience cohort were involved in site visits to either the Smith Family Centre or Apollo House where they worked for an hour a week and mostly one-to-one with school students who had been identified as having learning difficulties. Apollo House is an Indigenous neighbourhood centre run by the local women’s group. Funding was sourced from different organisations and the House mainly provides free services or links to other service providers to work on education, health and welfare. The Smith Family is a national organisation that supports students with learning needs through scholarships, peer and community support. It receives funding from a number of different sources (both local and state).

Consequently, PSTs from the applied experience cohort developed field site lesson plans and reviewed testing available in field sites. The lesson plans developed followed the structure of pedagogies studied by the embedded design cohort and included explicit teaching, cooperative learning, and differentiation of each. Each week during tutorial times PSTs from the applied experience cohort would develop lesson plans, and various assessments would be analysed for potential usefulness in their designated site. Examples of these assessments schemes included: running records for reading fluency and accuracy; South Australian Spelling Test (Westwood, 1999); and the Macquarie University MUSEC phonemic awareness tests (MUSEC, 2000). Examples of the cognitive strategies discussed included: phonological and morphological awareness cognitive strategies; and spelling strategies such as look, cover, write and check. The use of these strategies was modelled by the lecturer and the PSTs also had access to guided practice so the skills might be used on field visits. Use of all of these assessment approaches was not compulsory for PSTs in the applied experience cohort, rather they were used as appropriate by the PSTs.

PSTs in the applied experience cohort discussed the assessment strategies with peers, but not in the same structured group process or product format that had occurred in the collaborative groups from the embedded design cohort. Discussions were structured
based on the stage groups or year levels of the students the PSTs were working with. Discussion was framed initially in pairs, then in groups of four, and finally a class discussion was held about one important element that peers had given feedback on during the discussion time. Rubrics were provided for the lessons to be used in the field however these were not necessarily shared with peers in finished formats.

3.9.6 Teaching cycle for the applied experience cohort.
The primary education degree PSTs in the applied experience cohort undertook the inclusive classroom support experience. Lectures for the applied experience cohort followed exactly the same pattern and standard format as for the embedded design cohort. Additional information was presented in lecture format during week 13 on topics such as social and emotional issues students may face, and educational psychology theories (e.g., Theory of Mind) that could help PSTs better understand potential emotional difficulties the school students may face.

During the weeks where no lecture was held for the embedded design cohort, PSTs in the applied experience cohort were given additional input about the site visits they were engaging in. During this time, PSTs in the embedded design cohort were sharing draft lesson plans in collaborative groups for feedback in order to make changes prior to submission the following week. During workshop sessions for applied experience PSTs, additional information about the sites was disseminated, ranging from an introduction to the sites in week 1 to information about a variety of assessments that may be useful in-situ by week 4. Beginning in week 4, PSTs in the applied experience cohort would apply their lesson plans in their direct experience settings, with feedback provided by the instructor who monitored the process. Feedback from peers was given in tutorials the following week. Day book lesson plans (i.e., brief entries about content to be covered) were kept rather than the formal lesson outlines that were utilised by the embedded design cohort. Each week, a variety of assessment practices and cognitive strategies were discussed as appropriate and being used in the field. The strategy of focus was also incorporated during tutorial lesson design time for applied experience participants, with feedback given initially from a peer, then a group of four, then summary information presented to the class at the conclusion of the workshop.

PSTs in the applied experience cohort completed the same tests and quizzes as those in the embedded design cohort. These were scheduled in weeks 4, 8 and 12. PSTs in the
applied experience cohort submitted a folio of lessons developed over the course which constituted assessment item 2 (see Appendix 2). When composing these lessons, PSTs in the applied experience cohort were required to utilise the pedagogies in the course (explicit teaching and cooperative learning). Those in the applied experience design completed a summary paper for submission of assessment 2. The paper was of 3,000 words compiled in report form, was informed by the notes they had kept in their weekly journals, and covered areas such as:

- A general introduction to the child
- Identification of the basic learning needs of the child
- Short weekly lesson plans (day book format) and evaluations
- Weekly monitoring data
- Reflective and critical analysis of learning
- Examination of findings in relation to the research literature.

The PSTs in the applied experience cohort then participated in an additional one hour per week inclusive classroom support experience scheduled throughout weeks 3-13. Students worked with individual school students. Activities centred on literacy and numeracy skills including: guided reading, home reading, small group activities, and one-to-one guided practice with literacy skills. The groups included school students who had a documented special educational need as well as their peers. The lecturer from the applied experience cohort monitored site visits as Apollo House was scheduled from 3.30-4.30 pm and The Smith Family from 4-5 pm. Assistance was also provided in supervision by the field managers from each site. This activity translated to assessment item two for PSTs enrolled in the applied experience cohort (detailed above).

A summary of the differences between the embedded design and the applied experience cohorts is presented in Table 3.8.

<table>
<thead>
<tr>
<th>Point of comparison</th>
<th>Embedded design</th>
<th>Applied experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>85</td>
<td>24</td>
</tr>
<tr>
<td>Allocated time</td>
<td>Total of 39 contact hours</td>
<td>Total of 50 hours. 39 contact hours as per embedded design and 11 additional hours in</td>
</tr>
</tbody>
</table>
settings working directly with school students.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Embedded design components:</th>
<th>Embedded design components:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level 1: Knowledge and</td>
<td>Level 1 Knowledge and</td>
</tr>
<tr>
<td></td>
<td>awareness.</td>
<td>awareness.</td>
</tr>
<tr>
<td></td>
<td>Level 2: Active experience</td>
<td>Level 2: Active experience.</td>
</tr>
<tr>
<td></td>
<td>Level 3: Continuous feedback</td>
<td>Level 3: Feedback given</td>
</tr>
<tr>
<td></td>
<td>on lesson plans and quiz prep.</td>
<td>following in-situ meetings</td>
</tr>
<tr>
<td></td>
<td>Level 4: Personal impact via</td>
<td>with students.</td>
</tr>
<tr>
<td></td>
<td>quiz prep.</td>
<td>Level 4: Neither quiz prep</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nor lesson feedback was</td>
</tr>
<tr>
<td></td>
<td></td>
<td>structured in a collaborative way.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lecture and tutorial attendance</th>
<th>8 hour-long lectures and 13 x 2 hour tutorials. 6 x hour-long quiz preparation times in lieu of lecture (See appendix 3).</th>
<th>8 hour-long lectures and 10 x 2 hour tutorials. 6 x hour-long meetings to prepare for in-situ visits. 11 hours of visits working 1 on 1 with school students (See appendix 3).</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Course objectives covered</th>
<th>Topics in appendix 3 to elaborate</th>
<th>Topics in appendix 3 to elaborate.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Assessment items</th>
<th>Quizzes</th>
<th>Quizzes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lesson plans based on in-class tutorial experiences, written up using a formal template of pedagogical characteristics.</td>
<td>Lesson plans based on in-situ experiences with school students, written up in essay format.</td>
</tr>
</tbody>
</table>

3.9.7 Reliability of instructor implementation

In order to address the potential differences in implementation between three instructors, an implementation fidelity scoring protocol was constructed by the researcher (see appendix 1). Each week prior to lectures and tutorials, a group meeting was held between the three instructors to go over the content to be covered. At the conclusion of each week, the instructors filled in the scoring protocol and emailed results to the researcher. These were then summarised into the table found in Appendix 1. Each week the researcher sent all PowerPoints used in lecture delivery and activities planned to be covered in tutorials to the corresponding instructors of the course. Towards the end of the week, the researcher emailed a table document that summarised these activities and content to the coordinators where a scale from the protocol was to be completed ranging from:

- Strong agreement = (yes I did all of this during the week).
- Moderate = (I did some of this during the week).
- Weak = (I did not do this during the week).
The scale was a measure of how individual instructors felt they covered the content at each site. These findings are reported in the results chapter (Chapter 4). Where there was moderate or weak agreement, further details were sought and recorded by the researcher. These were also noted in Appendix 1.

3.10 Data Collection Procedures

Prior to data collection for study one, an ethics application was required from the University. Once obtained (Appendix 10), demographic information was sought from volunteers enrolled in the pre-service teacher education courses across the embedded design and applied experience cohorts. The pre-test for pedagogical content knowledge (PCK) was administered as well as self-efficacy (SE) during week one tutorial time across cohorts embedded design and applied experience by the principal researcher (See appendix 3). Further criterion-based mastery quizzes comprising five open-ended questions were also administered throughout the session at weeks 4, 8 and 12 to test the pedagogical content most recently covered. Every occasion for a quiz was also used to collect reflection data from participants. At the conclusion of each quiz, participants were required to write for 30 minutes about their learning on the most recent topic. These reflections written following the quizzes were used to check frequency and the sophistication of pattern language.

Self-efficacy questionnaires (both pre and post) were identical across cohorts in form and timing of delivery. Coding of the self-efficacy scores, the pre and post PCK tests, and the analysis of the reflections all occurred at the conclusion of the course after final grades had been submitted to the university. The quizzes constituted part of the grading for the course so these were marked and mastery determined for the students concerned.

During study one of this research, three data collection instruments were employed to gather information from the participants about their progress on the dependent variables of interest. Study one used the ‘Self-efficacy Towards Interactions with Students who have a Disability’ (SEIPD) questionnaire (Hickson, 1995) to determine levels of self-efficacy when working with students who have disabilities; researcher-developed tests and quizzes of PCK; and a reflection that was analysed for frequency
and level of sophistication of professional language used. Figure 3.3 documents the sequence of data collection for study one.

![Data collection time frames for study one.](image)

Figure 3.3. Data collection time frames for study one.

Data were collected from participants on four occasions: the first was prior to commencement of the intervention, second and third occasions were during session, and the fourth took place following the intervention period (one university semester of about 6 months). On receipt of the pre occasion data, each participant was allocated a case number in order to match the pre occasion data with the post occasion data. All data were amalgamated into SPSS according to individual case numbers and demographics to create a complete data set.

### 3.10.1 Dependent measures used for study one.

Study one dependent variables included the degree of self-efficacy experienced by pre-service teachers; the degree of pedagogical content knowledge understood; and their use of pattern language and the sophistication of that language to discuss pedagogies. Study one dependent variables were seen as the capacity building tools needed by PSTs in order to work more efficiently with students who have differing needs.

The dependent measures are used to evaluate change in the dependent variables and need to take into account issues of reliability and validity in order to indicate the authenticity of the data collected. Reliability is related to the consistency of the data over repeated measures. Validity considered whether the intended constructs were actually being measured (Kervin et al., 2006); or with appropriate interpretations made from scores obtained (Gay & Airasian, 2003). Validity is a matter of degree and is specific to interpretations being made for the groups being tested (Gay & Airasian, 2003). Details for each of these instruments are now discussed.
3.10.1.1 Self-Efficacy questionnaire (SEIPD).

The Self-Efficacy toward future Interactions with People with Disabilities scale (SEIPD) (Hickson, 1995) was a standardised instrument employed in study one and two. Authors such as Gay and Airasian (2003, pp. 147, 157) and Johnson and Christensen (2008, p. 163) suggest that researchers select a group of tests that might be appropriate and compare them on relevant factors before selecting the most suited. A cluster of factors were compared such as reliability and validity, cost of administration, time to complete, group or individual administration, and guidelines for use. The SEIPD was selected for this study as at that time, no suitable scale had been developed that targeted inclusive education and PSTs. Other scales have been developed to test self-efficacy of teachers who are already in classrooms and also teacher attitudes (Chan, 2008; Tschannen-Moran et al., 1998). However, the SEIPD was the closest scale available for PSTs even though it still had a medical slant to it as it included undergraduate nurses in the sample population. The SEIPD had also shown improvements for participants following intervention in an earlier study using PSTs (see Lancaster & Bain, 2007).

The scale was comprised of 15 items in three areas: willingness to initiate behaviour, willingness to expend effort in completing behaviour, and persistence in the face of adversity (Hickson, 1995). The questionnaire was structured in format and comprised closed questions in order to focus on a rating scale response. The resulting data was ordinal and needed to be treated with non-parametric analysis (Cohen et al., 2007). The SEIPD employs a Likert type 8–point scale ranging from “definitely false, false, mostly false, more false than true, more true than false, mostly true, true, definitely true” with no midpoint as a format for responding. For example: “I am able to plan and organize appropriate activities for my students” (Hickson, 1995, p.111).

3.10.1.2 Reliability of the SEIPD.

The reliability of the SEIPD was determined using test-retest and alpha coefficients employing a sample of 180 pre-service teachers and nurses (Gay & Airasian, 2003). A mean alpha co-efficient of .87 was reported for the SEIPD, while test-retest reliability produced a reliability coefficient of .8 over a four-week interval and .68 over a six-week interval (Hickson, 1995). Factorial validity was established using principal component analysis. Both orthogonal and oblique rotations gave identical results with only one factor extracted, indicating that items within the scale were measuring the
same construct and accounting for an average of 55.1% of the variance (Hickson, 1995).

3.10.1.3 Validity of SEIPD.

Construct validity refers to those constructs actually being measured by the test (Burns, 2000; Cohen et al., 2007; Gay & Airasian, 2003). Statistically significant correlations (.32) were found between the SEIPD and the Scale of Attitudes Toward Disabled Persons (Antonak, 1979, cited in Hickson, 1995) and the Attitudes Toward People with Disabilities Scale (.50) (Yuker, Block, & Young, 1970, cited in Hickson, 1995), confirming the mediating relationship of attitude on self-efficacy. As the authors note, no single validation study can establish construct validity of a measure (Gay & Airasian, 2003, p. 140). These measures of validity computed by Hickson (1995) also established discriminant validity between the constructs of self-efficacy and that of attitude (Shaughnessy et al., 2009, p. 164).

The original questionnaire was used to determine pre-service teaching and nursing student attitudes towards people who have a disability. The current use of the instrument was altered to include the word ‘student’ more frequently so it would hold more ‘face validity’ for the PSTs. For example: item 1 originally read “I feel confident in my ability to work with individuals with disabilities” was changed to “I feel confident in my ability to teach students with disabilities”. The scores achieved on this SEIPD questionnaire were not used to predict performance (Burns, 2000) but rather to establish the PSTs’ current level of self-efficacy of the construct and as a baseline measure to check individual change on the same construct at a later point in time.

The data collection tools for pedagogical content knowledge consisted of researcher made instruments. Lankshear and Knobel (2004, p. 163) discussed essential steps to ensure the quality of researcher made tools such as: identify the purpose and target population of the measure; determine the relevant behaviours to be measured; identify relevant importance of the selected behaviours; and investigate the reliability and validity of each.

In all instances, the target populations for this research were PSTs who were studying a mandatory course in inclusive education as part of their undergraduate teaching
degree. The purpose was specifically to determine if a particular course design, with specified elements from self-organizing theory, had an impact on dependent variables of interest relating to skills and attitudes towards students who have disabilities. The relevant behaviours to be measured thus incorporated attitudinal as well as skill-based behaviours. The importance of the selected behaviours was situated in the theory of self-efficacy and the skills were determined based on previous research into efficacious strategies such as explicit teaching (ET) and cooperative learning (CL). These pedagogies possessed an extensive research history that shows statistically robust effects in controlled experimental research over time, settings, and populations (e.g., Bloom, 1984; Fraser et al., 1987; Gillies & Ashman, 2000; Slavin, 1991).

3.10.2.1 Pedagogical content knowledge (PCK).

The pedagogical content knowledge (PCK) test was a criterion-referenced test that involved a comparison against predetermined levels of performance rather than against other students (Gay & Airasian, 2003). The PCK scores were derived from a set of 20 researcher-developed questions that framed the content to be covered throughout the course.

During weeks 1 and 12, a pre- and post-test were administered using the same questions across both conditions. Further, formative criterion-based mastery quizzes comprising five open-ended questions were also administered throughout the session (at weeks 4 and 8) to test the pedagogical content most recently covered. All quizzes consisted of 5 questions and attracted a total possible score of 12 marks. Quiz 1 covered explicit teaching, collaborative problem solving and introductory inclusive education principles. Questions included those at factual and comprehension levels (for example: Describe the research based characteristics of explicit teaching as a pedagogy. Which characteristics are the most important and why?) as well as at the application level (for example: Devise a learner outcome incorporating the use of cognitive strategies to differentiate the content in a HSIE lesson). Opportunity was also given for questions at the synthesis level (for example: How do the characteristics of cooperative learning fit within the framework of Motivational theory?) (See Appendix 4 for example). Quiz 2 covered cognitive strategies as well as differentiation. Quiz 3 covered cooperative learning and adaptations that can be made. Five questions within the post-test that focused on these topics of cooperative learning and adaptations were situated within the post-test administered. The score for quiz 3 was calculated by
adding marks for these five questions. Marks from quizzes 1, 2 and 3 were counted towards the PST grade for the course. Quiz 3 became the post-test of the baseline measure administered in week one. The entire post-test of 20 questions was not counted towards the overall grade for the course but was scored and used for the pre/post comparison.

3.10.2.3 Content validity of the pedagogical content knowledge test.

Researchers are concerned with the content validity of a test as it is the representativeness or sampling adequacy of the content (Burns, 2000, p. 351; Cohen et al., 2007). It is achieved by making personal judgements about the sampling of a particular domain; in this case, research-based inclusive pedagogies. It is concerned with how well the items (content) in the tests and quizzes represented PCK of inclusive education pedagogies, and how well the items measured PST mastery of these concepts. Some suggest this is indicative of item validity (Gay & Airasian, 2003, p. 136). While it does not claim to represent the ‘universe’ of the content being measured, it does represent the universe of the content set out in the learning objectives for the course, or sampling validity (Gay & Airasian, 2003). The content validity of the ‘pedagogical knowledge test’ and quizzes was enhanced as the test fairly represented the actual content that was taught across the weeks of the session. In the inclusive education course this was evidenced as reading objectives given to PSTs each week. These course objectives were made the focus of lecture and tutorial presentations/activities and the same objectives where questions were framed as questions used in the quizzes. The test measured what the PSTs were expected to learn. This repetition was part of the embedded design. In this way, the content validity was representative of the sampling adequacy.

Content validity is often determined by expert judgement (Cohen et al., 2007; Gay & Airasian, 2003). In this instance, the entire content covered in the course was determined by the inclusive education team (of 5 members) as part of the course redesign. Ideas of best practice were sourced from group members based squarely on the principle elaborated above of ‘simple rules’. Simple rules might look like the following: learning is cooperative, decision-making is collaborative; and practice is based on research. The simple rules were delineated in each of the course assessment items so PSTs could see the rationale behind the course structure. They were made
explicit for the PSTs in an ongoing manner and continually revisited throughout the session.

The warning advanced by Burns (2000, p. 352) is heeded: “one should keep in mind that content validity of the test is not a fixed and changeless characteristic”. Thus, the tests were made up especially for the groups to be tested and based on the exact content on which they were to be tested; validity is not a property of the ‘instrument’ but rather the property based on measuring particular participants in a particular context (Johnson & Christensen, 2008, p. 151).

3.10.2.2 Reliability of pedagogical content knowledge.

There were a number of factors that could affect reliability identified by Cohen et al., (2007, pp. 159-163), and a range of moderation strategies were used both before and after the tests to enhance the reliability of researcher-made tests: conducting group moderation of grades, making post hoc adjustments to grades, defining marking criteria, providing exemplification, and holding group moderation meetings. Johnson and Christensen (2008) agreed with the recommendation that the reliability of instruments be determined empirically and suggested four main ways of computing indexes of reliability: test-retest, equivalent forms, internal consistency, and inter-scorer. The authors suggest that several different ways of computing reliability may be useful to provide corroborating evidence. These will now be elaborated in terms of the current study.

Agreement trials and practice marking for interscorer reliability were all undertaken immediately following each PCK test or quiz (see the results chapter, Chapter 4, for details of inter-rater reliability obtained). Reliability of marking was achieved between the researcher, colleagues and a research volunteer by jointly grading 20% of marking, randomly selected tests and quizzes across tutorials and campuses. A decision table was constructed to facilitate coding and to ensure a high degree of consistency (Johnson & Christensen, 2008). Discussions were held after 5 to 10 scoring attempts had been completed to minimise any confusion and to attain a reliability score of greater than 90%. This was seen to be even more important when there were large numbers of papers and multiple markers (Cohen et al., 2007).
It was also possible to address reliability through calculation of reliability coefficients such as *split–half* (applied to pre- and post-tests). These reliabilities are reported in the results chapter. As there were 20 items in each, the Kuder-Richardson formula was also applied. The tests were framed as *mastery or criterion-referenced*. There was an expectation the results would hover around the high level and thus the range was limited. This in turn lowered the correlation coefficients that may be calculated (Burns, 2000; Cohen et al., 2007; Gay & Airasian, 2003). The raw scores were more beneficial here for these calculations. Reliability and validity of data collection formats were both seen as essential properties to consider. One of the problems associated with use of the split-half method of calculation is that different results may be obtained from different ways of subdividing the test (Johnson & Christensen, 2008).

With this in mind, a coefficient alpha may be a better measure to calculate a correlation coefficient. Coefficient alpha is an estimate that averages *all possible* split-half correlations, corrected by the Spearman-Brown formula (Johnson & Christensen, 2008) that recommends a result of 0.7 or higher for research purposes to determine if items are homogeneous or internally consistent, and as high as > 0.9 for clinical use. These authors warn that high numbers of items will automatically favour a high correlation score. In this instance, 20 items would not be deemed to cause such a result.

Reliability was seen as a necessary but not sufficient condition for validity as it is no guarantee of validity (Johnson & Christensen, 2008). Validity will be discussed in terms of content and construct validity issues.

3.10.2.4 *Construct validity of the pedagogical content knowledge test.*

This validity is concerned with whether the test items were indicators of the underlying construct in question. Underrepresentation is a danger where the test becomes too narrow. In the instance of *embedded design*, this threat was minimised as the construct was studied in a number of ways including knowledge quizzes, lesson designs and differentiation, and the use of pattern language to discuss the pedagogy characteristics. Members of the inclusive education team agreed on important operationalisation of the construct. The definitions were confirmed through literature reviews of explicit teaching and cooperative learning research based pedagogies used in the field.
As a researcher made instrument, it was not appropriate to compare results to any other standardised formats available. Johnson and Christensen (2008) discuss the possibility of using validity evidence based on internal structures of the test such as factor analysis or measures of test homogeneity (item-to-total correlation). If coefficients of alpha are low (i.e., less than 0.7), it might suggest some items are measuring different constructs. When used to discuss validity, the correlation coefficients were referred to as validity coefficients.

3.10.3.1 Pattern language (PL)

Pattern language was measured by asking PSTs to write a reflection about their lesson designs on three occasions throughout the course. As closed questions were thought to potentially limit the sophistication of the responses in a previous study (Bain, Lancaster & Zundans, 2009), these were not incorporated in this instance. PSTs were asked to write for up to 30 minutes about their learning of the most recent pedagogy they had been studying. Writing a reflection consisted of responding to an open ended item rather than posing specific questions to be answered. For example: “Write for 30 minutes about your learning”. Reflections were written after the completion of each teaching cycle in weeks 4, 8 and 12.

The PSTs were asked to draw upon their own experiences in developing the lesson designs and the feedback they received from their peers. Also differing from the previous study (Bain, Lancaster & Zundans, 2009), PSTs were encouraged to try and include pattern language in their reflections. The reflections were allocated two marks as part of the quiz grade in study one: frequency and sophistication of pattern language. The reflection questions did not require PSTs to incorporate knowledge from prior reflections although it was possible to do so if they chose. The data from these reflection samples were required for two different dimensions of pattern language: one of frequency and one of sophistication.

3.10.3.2 Frequency of pattern language.

The researcher and a research volunteer were responsible for coding the reflections completed at the end of each quiz. Again, following the use of the embedded design principle, the content and topics that had been the focus of instruction week by week, and also the focus of the assessment quiz, were listed as the items to locate for a frequency count in the reflection. For example, the vocabulary associated with explicit
teaching were noted by the researcher and located within the narrative of the reflection. The same procedures were followed for cooperative learning. Table 3.9 is an indicator of terms scorers would take into account.

**Table 3.9**

*Pattern Language Terms*

<table>
<thead>
<tr>
<th>Explicit teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Outcomes/objectives identified (who, what, criteria)</td>
</tr>
<tr>
<td>• Anticipatory set</td>
</tr>
<tr>
<td>• Link to prior learning</td>
</tr>
<tr>
<td>• Teacher model</td>
</tr>
<tr>
<td>• Guided practice</td>
</tr>
<tr>
<td>• Independent practice</td>
</tr>
<tr>
<td>• ET in conjunction with mastery learning</td>
</tr>
<tr>
<td>• Differentiation enabled</td>
</tr>
<tr>
<td>• Top five characteristics of effective instruction are incorporated in ET if structured correctly: structure, sequence, reinforcement, practice and feedback</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cooperative learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Face-to-face interaction</td>
</tr>
<tr>
<td>• Positive interdependence</td>
</tr>
<tr>
<td>• Interpersonal skills</td>
</tr>
<tr>
<td>• Focus on group process</td>
</tr>
<tr>
<td>• Individual accountability</td>
</tr>
<tr>
<td>• Social cohesion</td>
</tr>
<tr>
<td>• Cognitive elaboration</td>
</tr>
<tr>
<td>• Metacognition</td>
</tr>
<tr>
<td>• Declarative</td>
</tr>
<tr>
<td>• Differentiation</td>
</tr>
</tbody>
</table>

These terms were used for coding the *frequency* of PL terms and were simply counted in each reflection. Terms were only counted once and the total number of different terms recorded for each participant.

Pattern language assessment also included the language terms of *cognitive strategies* as these were part of the curriculum being studied by PSTs in their inclusive education course. Cognitive strategies have an associated professional language and were included in the assessment of pattern language frequency and sophistication for this reason. The focus in this study, however, was on pedagogical approaches that could be observed in classrooms. It was thought that teaching of cognitive strategies by PSTs would be evident in either *explicit teaching* or *cooperative learning* lessons.
3.10.3.3 Sophistication of pattern language.

Once the data was coded for frequency of PL terms, it was revisited to analyse the complexity of the responses. The coding of the sophistication of responses was carried out using SOLO taxonomy originally developed by Biggs and Collis (1982). Sophistication is described as the level of cognitive responses given by participants in their reflections. The levels ranged from a very simple response where the PSTs provided one piece of information to the very complex where pieces of information were presented that linked information together and also included explanations as to the theory or research behind those links. The following table (Table 3.10) will elaborate the criteria required for each category of SOLO.

*Table 3.10*

SOLO Coding System Categories (Biggs & Collis, 1982)

<table>
<thead>
<tr>
<th>Code</th>
<th>SOLO Level</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Blank</td>
<td>The explanation section has been left blank and no explanation is provided.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The response does not appear to answer the question or may simply be stating the question.</td>
</tr>
<tr>
<td>1</td>
<td>Prestructural</td>
<td>One piece of information was evident in the response.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Responses at this level contain one fact. More than one piece of information was provided in the explanation. Responses at this level contain several facts, but consider the facts in isolation; no clear links are made amongst the facts.</td>
</tr>
<tr>
<td>2</td>
<td>Unistructural</td>
<td>Pieces of information have been presented and related together. Various facts are linked together and are related to a main concept; the explanation is valid only for the given context.</td>
</tr>
<tr>
<td>3</td>
<td>Multistructural</td>
<td>A response of this type goes beyond what is asked in the question however the explanation presented by the respondent clearly indicates how the additional information relates to the question. The response generalises across contexts.</td>
</tr>
<tr>
<td>4</td>
<td>Relational</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Extended Abstract</td>
<td></td>
</tr>
</tbody>
</table>
3.10.3.4 Reliability of pattern language frequency and sophistication.
Biggs and Collis (1982) discuss issues of reliability for the SOLO classification system in detail. It is important to note that SOLO is a *response* measure and not a measure of personal characteristic. As such it is sensitive to instruction, and for this reason, the most appropriate measure to check reliability is inter-judge agreement. Agreement was sought from these authors in areas of history, poetry and creative writing. The correlations between the two judges following analysis of a ‘historical letter’ was 0.81 (N = 107) and 0.71 (N = 104); in cases of creative writing the scores were 0.79 (N = 63) and .83 (N = 51); and poetry .95 (N = 26) and .78 (N = 26).

This same inter-rater measure of reliability was used in the current research and was achieved by jointly coding 20% of all reflections from each quiz between researcher, colleagues and a research volunteer. This occurred for *frequency of PL* terms and *SOLO level* for both study one and study two. A decision table was constructed to facilitate coding and to ensure a high degree of consistency. The decision table is described in Appendix 6. Discussions were held after 5 to 10 quizzes had been marked to minimise any confusion and to attain a reliability score of greater than 90%.

3.10.3.5 Validity of pattern language.
In order to determine validity of the SOLO measure, Biggs and Collis (1982, p. 189) used teacher ratings and factor analysis procedures. They sought 100 responses to two poems and 100 essays that were selected at random from Grade 9 students. With an experienced teacher rating them on whatever scale thought to be most appropriate, agreement was not as high as the inter-judge correlations but was still reported as “reasonable” (p. 190). Factor analysis was employed in 1981 by Kirby and Biggs (cited in Biggs & Collis, 1982) using 16 known tests on cognitive abilities, school achievement, motives for learning and learning strategies. Results indicated that SOLO related to school achievement, in the way teachers currently and typically assess students. Factor analysis showed appropriate loadings of SOLO on items of school achievement. Process analysis was used with canonical correlation to relate overall SOLO ratings and transitions from level to level within SOLO, and with appropriate indices of cognitive processes, student motivation and learning strategies.
To investigate the nature of processes used, in 1994, Boulton-Lewis conducted a study with a random sample of 100 students across courses and faculties. They were asked to define ‘learning’ and their responses were coded using SOLO categories. A one-way MANOVA for three SOLO levels (unistructural, multistructural, and relational) was conducted with deep, surface and achieving motives and strategies (measured using the Study Process Questionnaire [SPQ]). The result was significant (p = .03), demonstrating convergent validity with the SPQ. The results indicated a decline in surface motivation and strategies used by students with an increase in structural organisation, as measured by SOLO. Tukey post hoc analyses showed significant differences between deep motive and deep strategies and between unistructural and relational SOLO levels (Boulton-Lewis, 1994).

Chan, Tsui, Chan, and Hong (2002) empirically explored the application of three different taxonomies in measuring cognitive learning outcomes: SOLO taxonomy, Bloom’s taxonomy and a reflective thinking model were compared. The findings indicated that SOLO was suitable for measuring different kinds of learning outcomes, however, conceptual ambiguity was persistent amongst the higher categories. Sub-levels of SOLO were added for analysis as well from studies conducted by Trigwell and Poser (1991) and Burnett (1999; cited in Chan et al., 2002). All three taxonomies correlated with each other (r = 0.6 for SOLO; r = 0.93 for Bloom’s; and r = 0.87 for the reflective thinking model).

In the current research, content validity was concerned with how well the items (content) in the reflections represented the professional pattern language of inclusive education pedagogies, and how well the items measured PST mastery of these concepts. Content validity is often determined by expert judgement (as noted in the investigation above by Biggs & Collis, 1982). In this instance, the entire content covered in the course was determined by the inclusive education team as part of the course redesign. Ideas of best practice were sourced from group members based squarely on the principle elaborated above of ‘simple rules’. Simple rules might look like the following: learning is cooperative, decision-making is collaborative; and practice is based on research. The simple rules are delineated in each of the subject assessment items so PSTs could see the rationale behind course structure. They were made explicit for the PSTs in an ongoing manner and continually revisited throughout the session.
3.11 SI Data Analysis

A number of data analytic techniques were used to determine the effect of the experimental condition. Once the data were collected, statistical analysis was used to summarise and interpret the findings. The two main branches of statistical analysis that were used included: descriptive and inferential statistics. Descriptive statistics were deployed as a means of organising collected data while inferential statistics were used in study one to draw causal inferences. One common statistical method used to separate the effect of the treatment from the effect of selection differences is the use of ANOVA (analysis of variance). ANOVA was a key analytical technique used in this study. It specifies three components of post data response: the grand mean across all participants; the treatment effect which is the average the treatment adds; and the error representing all other factors that may influence the results. The pre-test scores were also taken into account in this research so more concise definitions of group comparability could be made at the same time. The caution by Burns (2000) is heeded where even though mean pre-test differences may be small between groups; this is not necessarily an indication of equivalence. A series of two-way repeated measures ANOVA were conducted for each set of measures. The intervention condition (embedded design v applied experience) was used as a between subjects factor and time as a within subjects factor. One-way repeated measures ANOVA and paired samples t-tests were conducted to follow up on statistically significant main effects. Holm’s sequential Bonferroni procedure was used to control for Type 1 error across pairwise comparisons (α = .05).

3.11.1 Pedagogical content knowledge analysis.

A one-way repeated measures ANOVA was conducted with intervention condition (embedded design versus applied experience) as a between subject’s factor and time (pre-test versus post-test) as a within subjects factor (Allen & Bennett, 2010, p. 114). The Condition x Time interaction was examined using the multivariate criterion of Wilks’s Lambda (Λ). T-tests were used as a simple effects analysis to follow on from a statistically significant interaction. Holm’s sequential Bonferroni procedure was used to control for Type 1 error across comparisons for simple effects with alpha set at .01.

Pedagogical Content Knowledge (PCK) was based on scores obtained for quizzes 1, 2 and the final 3. Quizzes 1 and 2 consisted of five questions each in a similar format. In
quiz 3, only the five questions pertaining to the new content of cooperative learning and differentiation were calculated in the score recorded. In order to compare pre- and post-score differences, the complete baseline quiz and the complete quiz 3 were compared. These consisted of 20 questions that covered all content knowledge covered in the course.

Analysis of PCK was enabled by the coding process which was conducted by the researcher and a colleague teaching tutorials during the course. Answer sheets were devised to correspond with each of the scores out of a total of /12, reported for quizzes 1, 2 and 3. Scores out of a total of /35 were reported for the pre- and post-tests across conditions. Reliability of scoring was checked by taking 20% of each of the testing occasions and double marking clean copies. Greater than 90% reliability was achieved for each occasion by noting agreements divided by agreements plus disagreements between the researcher, other teaching staff, and research assistant (see Appendix 4).

3.11.2 Pattern language frequency.
The PST reflections were analysed in two ways. First, a frequency count was taken of the number of pattern language terms included in each PST response. Pattern language terms were defined as those words that comprised the professional lexicon of the teaching approach or strategy taught in the class. For example, with respect to explicit teaching, words like *modelling guided practice, anticipatory set, independent practice*. These terms describe the critical sub-components of the pedagogies, knowledge of which is essential to implement the pedagogies with integrity and to problem solve their use in classroom settings.

3.11.3 Pattern language sophistication.
The ways in which the terms were used in the narrative constituted the second form of data analysis. It was considered possible to use pattern language terms frequently as part of a reflection in ways that did not necessarily have clear meaning, communicative intent, or show any level of sophistication in understanding or analysis. It is also possible that a sophisticated response could be produced without pattern language terms, although the successful use of the practices included in the study is predicated upon knowledge of their structural elements (e.g., task structure, guided practice). The use of terminology was considered and related to those structural elements to be an important component of a sophisticated response. The Structure of Observed Learning
Outcomes taxonomy (SOLO, Biggs & Collis, 1982) was used to make a determination of the sophistication of the reflection narratives and to address the way pattern language terms were used.

The SOLO was developed by Biggs and Collis (1982) as a means of assessing the sophistication of learner responses across a range of domains and across students of various ages (Chan et al., 2002). The taxonomy was structured into five major levels and is hierarchical in nature, increasing in structural complexity. These hierarchical levels reflect the quality of learning for a particular task and are suited to the content analysis of prose passages or process analysis such as mathematical problem solving (Biggs, 1995). SOLO has been used extensively in assessing responses including secondary science (Levins, 1997); knowledge of biology, in particular evolution amongst Stage 6 students (Creedy, 1993); use of LOGO computer language (Hawkins & Hedberg, 1986); the visual arts in higher education (Hulsbosch, 2006); and assessment in higher education across subject areas (Biggs, 1992).

In the present study, a trained research volunteer who did not possess knowledge of the study’s research questions undertook joint coding and analysis of the reflections. In the first round of analysis, the assistant identified all instances of use of the pattern language terms on each of the reflections. In the second round of analysis, each reflection was reviewed and coded according to the SOLO level to which it corresponded. The identification of terms of the assistant were compared to ratings made by the researcher for 20% of the reflections. The checks achieved or exceeded 90% agreement for the identification of terms and the designations of response sophistication on the SOLO taxonomy. Reliability was calculated by determining the instances of coding agreements for both factors in the reflections across the sample for the two raters and then dividing those by coding agreements plus disagreements. This included agreements/disagreements for the presence of pattern language terms and the SOLO level of coded responses.

Samples of data for pattern language were collected in exactly the same manner across conditions. The three lecturers were in constant contact in terms of language that was to be used with the PSTs, and timeframes for completions, etc. The lecturer from the applied experience cohort sent the documents to the researcher for analysis.
Quiz 1 reflections covered topics of: inclusion, collaborative problem solving (CPS), and explicit teaching (ET). As there were three ‘topics’ covered, a SOLO score was designated for each of the three topics as well as an average score for the ‘quiz occasion’. Quiz 2 topics consisted of cognitive strategies as a way of differentiating ET. A SOLO score out of 5 was allocated. To score 3 the answer needed to include multiple entries of ‘cognitive strategy’ points and a score of 4 needed to involve linking between the ‘cognitive strategy’ points. If ‘explicit teaching’ and ‘cognitive strategy’ were linked this might be considered to be at level 4. If theory had been incorporated it would be scored at level 5. To score 5 the PST needed to link with theory and research and the strategy of focus. To score 4 the PST needed to relate areas of ‘cognitive strategy’ and ‘explicit teaching’ as well as relating points to each other in each area. To score 3 participants needed points in both areas not just one.

The focus in quiz 3 was Cooperative Learning (CL) and differentiation of CL (DiffCL). This topic was awarded the SOLO level of 5 with the usual requirements of multiple points noted scoring a /3; linking of these points rating a 4; and linking to theory and/or research rating 5.

Overall content of this quiz was the accumulation of the subject so it was possible to have mentioned points from all previous topics covered. A SOLO score of 5 was allocated to use of overarching theme/concept of inclusion/variety of differing needs; variety of differing students, etc, to the use of topics covered such as collaborative problem solving, explicit teaching, differentiation, and cooperative learning. A score of 4 meant there was a connection made between the points listed for discussion of the topics (i.e., linking discussions between explicit teaching and differentiation; or linking explicit teaching and the use of cognitive strategies). A score of 5 would have been designated if these were then linked to research or theoretical underpinnings (see Appendix 5).

Sophistication of pattern language was analysed using a Chi-square. Contingency coefficient C was used for this analysis as the variables are in the form of categories. The Chi-square of independence of categorical variables is most commonly carried out with the use of a contingency table (Burns, 2000).
3.11.4 SEIPD analysis.

In study one, Repeated Measures ANOVA was used to test mean differences between intervention conditions for occasion (pre to post) and if the differences were due to chance or the systematic treatment effects. This was deemed essential given that students could not be assigned randomly to the two conditions. The Condition x Time interaction was examined using the multivariate criterion of Wilks’s Lambda (Λ). A one-way repeated measures ANOVA was used to follow up a statistically significant main effect. Holm’s sequential Bonferroni procedure was used to control for Type 1 error across comparisons for simple effects that followed from the one way ANOVA. Alpha was set at the .05 level.

3.12 Study two

Details will now be provided for study two. This section of the chapter will:

- List research questions and hypotheses;
- Describe the research design employed to investigate the independent and dependent variables of interest;
- Discuss threats to validity of the method and how these were countered in the current research;
- Describe the procedures followed for study two, including the researcher’s access to participants, settings and ethical considerations;
- Describe levels of the independent variable for experimental and control groups;
- Describe the instrumentation process used to collect and code data and issues of validity and reliability of each; and finally
- Describe the data analysis procedures.

3.12.1 Research questions

The overarching research questions in study two are as follows:

Does application of the embedded design principle covary with improvements in:

- The design of inclusive education pedagogies used by PSTs in the professional experience setting?
- The implementation of inclusive education pedagogies used by PSTs in the professional experience setting?
• The self-reflection of PSTs about their lesson design and delivery?
• The frequency and sophistication of pattern language? And the
• Self-efficacy of PSTs?

3.13 Hypotheses

H₂ = Descriptive statistics for observation of classroom implementation will show higher levels of:

a) Design integrity for PSTs who have experienced embedded course design when implementing ET or CL compared to those who experienced the applied course design;

b) Implementation integrity for PSTs who have experienced embedded course design when implementing ET or CL compared to those who experienced the applied experience course design;

c) Self-awareness of PSTs about lesson design and implementation for PSTs who have experienced embedded course design when implementing ET or CL compared to those who experienced the applied experience course design;

d) The frequency and sophistication of pattern language for PSTs who have experienced embedded course design when implementing ET or CL compared to those who experienced the applied experience course design; and

e) Higher levels of self-efficacy between the PSTs who experience the application of embedded design principle to the theoretical design of the inclusive education course as opposed to those who participate in an applied experience condition.

H₀₂ = Descriptive statistics for the observation of classroom implementation will show there will be no demonstrable or meaningful differences between:

a) Design integrity for PSTs who have experienced embedded course design when implementing ET or CL compared to those who experienced the applied course design;

b) Implementation integrity for PSTs who have experienced embedded course design when implementing ET or CL compared to those who experienced the applied course design;
c) Self-awareness of PSTs about lesson design and implementation for PSTs who have experienced *embedded* course design when implementing ET or CL compared to those who experienced the *applied experience* course design;

d) The frequency and sophistication of pattern language for PSTs who have experienced *embedded* course design when implementing ET or CL compared to those who experienced the applied experience course design; and

e) Self-efficacy between the PSTs who experience the application of *embedded design* principle to the theoretical design of the inclusive education course as opposed to those who participate in an applied experience condition.

### 3.14 Research Design

The design for study two is presented in Figure 3.3. Study 2 investigated the ability of PSTs to design, translate, and then self-reflect on these skills into the fieldwork of practicum. A non-equivalent control group design was used in Study 2 as it was in Study 1 (see section 3.4).

*Figure 3.4. Study two: Non-equivalent control group design.*

Study two adopted multiple data collection methods to determine if research-based strategies had been maintained and generalised to the practical settings of schools 12 months after coursework, when PSTs were on practicum. Fraenkel and Wallen (2006)
used the design term ‘Static-Group Pre-test/Post-test Design’ (p. 272) which is represented as follows in terms of Study 2.

Table 3.11
The Static-Group Pre-test/Post-test Design

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>M</th>
<th>O₁</th>
<th>X₁</th>
<th>O₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>M</td>
<td>O₁</td>
<td>X₂</td>
<td>O₂</td>
</tr>
</tbody>
</table>

The dashed line refers to the two groups already formed. X₁ symbolises the experimental treatment (the embedded design independent variable). X₂ symbolises the applied experience independent variable. During study one (12 months earlier), the control group received a comparison condition which was seen as a more robust approach to compare the embedded design condition with an approach that has been identified in the literature as a source of benefit or efficacy by giving PSTs an experience in an applied setting as part of a course experience. O refers to the observations of both groups occurring at the same time. The subscripts of 1 and 2 denote the fact that the observations were the same in both settings in terms of time and content (Shaughnessy et al., 2009). It is important to note that O also represents several different co-variables of interest in study two. The changes following application of the independent variable were measured through use of pre-tests for the dependent variables of self-efficacy and pattern language. O₂ measured changes in the dependent variables of self-efficacy, use of pattern language, as well as unique dependent variables for study two of design, implementation and self-reflection fidelity of a lesson delivered on practicum.

3.14.1 Threats to validity and control over experimental designs

Study two was a continuation of study one, and as such shared a number of common elements (e.g., the overall design, use of the SEIPD and pattern language measures). However, differences in participants, procedures and dependent variables and measures called for separate descriptions of those elements. The following sections will elaborate on major threats to validity and control over the research design including: internal validity, statistical conclusion validity, and external validity.
3.15 Threats to Internal Validity

Eight common threats to internal validity: history, maturation, statistical regression, testing, instrumentation, selection bias, and attrition, as well as the interaction of these threats with each other identified by Burns (2000), Cook and Campbell (1979), Fraenkel and Wallen (2006), and Shadish, Cook, and Campbell (2002) are noted for their impact on study two in the following table.

Table 3.12
Potential Threats to Internal Validity in the Current Research

<table>
<thead>
<tr>
<th>Potential threats to internal validity</th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection Bias</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Maturation</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Attrition</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Additive and interaction effects (i.e., Selection – maturation interaction)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Instrumentation</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Statistical regression</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

3.15.1 Selection bias.

Selection bias is a major threat in group designs as it is impossible to randomly allocate participants to each of the groups (Shaughnessy et al., 2009). When random selection is not an option, as in the current study, groups can be selected that are as similar as possible and a pre-test of dependent measures administered to check for initial equivalence. To reduce the threat of selection biases in study two of this research, pre-occasion data were collected on self-efficacy and pattern language dependent variables of interest as well as demographic information in order to check for nuances between the groups. The results at pre-test level for self-efficacy, and use of professional pattern language indicated no significant differences and are reported in full in the Results Chapter (Chapter 4).
3.15.2 History.
Study two increased the likelihood of history having an impact on internal validity as the data collection was extended by a further 12 months. As a means to lessen the effect of this extended timeframe, the data collection points were held constant for both conditions. There was no real means of controlling for events of historical nature in this methodology so account of this was taken when discussing the results.

3.15.3 Maturation.
During the second study, a test of implementation fidelity occurred 18 months after the research was initiated, and 12 months after the post-testing for study one. Due to the placement in time of the course involved, and then the delay in practicum follow-up, issues of maturation could not be avoided for study two but rather controlled as much as the individual’s demographic and starting points can be compared from initial pre-test occasions. The threat of maturation to validity was also somewhat lessened as both groups were exposed to the same latency in testing occasions.

3.15.4 Attrition/Mortality.
These threats were considered as minor and addressed in the same way as study one. Refer to section 3.5.4

3.15.5 Additive and interactive effects to validity.
With the major threats of selection bias, history, maturation and attrition being identified in study two, it follows that the interactive effects of these will also be of major concern. The same precautions were observed as in study one. Refer to section 3.5.5. Subsequent coursework that PSTs completed in the interim timeframe was not measured and accounted for as part of this study. Both conditions were exposed to the same sequence of coursework as part of their programs. No controls over this were possible.

3.15.6 Testing.
In study two, there were no pre-tests involved for the lesson design and implementation, just the single data collection point, so this threat is not an issue in this instance. Testing for self-efficacy used the same questionnaire, and testing of pattern language incorporated the same measures as study one (see section 3.5.6).
3.15.7 *Instrumentation.*
These threats were considered as minor and were the same as for study one. Refer to section 3.5.7.

3.15.8 *Statistical regression.*
These threats were considered as minor and the same as for study one. Refer to section 3.5.8.

### 3.16 Threats to Statistical Conclusion Validity
Various types of statistical conclusion validity include: statistical power, violated assumptions of statistical tests, error rate issues, reliability of measures, reliability of instrument implementation, variance in the settings, and heterogeneity of participants.

<table>
<thead>
<tr>
<th></th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low statistical power</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Violated assumptions of statistical tests</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Error rate problem</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Reliability of measures</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Reliability of treatment implementation</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Extraneous variance in the experimentation setting</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity of respondents</td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

Low statistical power, violated assumptions of statistical tests, and error rate issues do not apply for study two. Inferential statistical tests were not used due to the small numbers of participants. The fact that the data is represented in descriptive form means that the benefit of inferential statistics to establish the probabilistic significance of an effect was not available and as such should be seen as a threat. The reader is left to make a judgement about the implications of any differences for the data provided.

Reliability of measures, reliability of instrument implementation, variance in the settings, and heterogeneity of participants all remained the same as in study one for self-efficacy (see section 3.10.1.1) and pattern language (see section 3.10.3.4). Study two also included an observation instrument to determine fidelity of: lesson designs, lesson implementation and self-reflection by PSTs of their lesson delivered in-situ.
The observation instrument developed for use in study 2 may be seen as ‘selective’ (Lankshear & Knobel, 2004, p. 221) as aspects of interest were identified and defined beforehand; that is, characteristics of explicit teaching and cooperative learning that are essential for the successful deployment of those pedagogies (Shaughnessy et al., 2009). Course observation protocols of high fidelity characteristics of explicit teaching and cooperative learning pedagogies were developed by the researcher. These characteristics were developed based on research literature and content covered during study one. (See Appendix 7 for the observation protocols used to assess PST design, implementation, and self-reflection fidelity). The threat to reliability of the data collection using the instrument in each of these instances was reduced by having a research assistant complete a sample of the protocols and inter-observer agreement reached above 90% on all occasions.

The observation instrument (the fidelity protocol) used in this study was used to gather data on three areas of PST work in school settings: PST design of their lessons, the implementation of the lesson, and also self-reflection about the same lesson following delivery. The implementation of the lesson was observed by the researcher and entries recorded on the protocol as the lesson progressed. The researcher used the same instrument to record entries about the lesson design fidelity (lesson plan) when provided by the PST. The PST was then asked to complete the same instrument in terms of a self-reflection about the lesson they had just delivered.

The variance in the experimental setting is not an issue in study two as no further instruction for the embedded design or the applied experience cohorts occurred. However, subsequent coursework that PSTs completed in the interim timeframe were not measured and accounted for as part of this study. Both conditions were exposed to the same sequence of coursework as part of their programs. No controls over this were possible. More concern is noted for the extraneous variance in the actual practicum setting as each participant was in a different setting and exposed to different supervising teachers. This threat was considered as major threat in study 2 due to the variations found in practicum settings.
3.17 Threats to Construct Validity

Threats to construct validity were all considered to be minor in study two and procedures were in place to further minimise these threats. The threats were the same as in study one (see section 3.7).

Table 3.14

<table>
<thead>
<tr>
<th>Potential Threats to Construct Validity</th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate operational explication of constructs</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>The Hawthorne Effect</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Experimenter effects</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Confounding constructs and levels of constructs</td>
<td>✔</td>
<td></td>
</tr>
</tbody>
</table>

Reliability testing was conducted for study two dependent measures using inter-rater agreement. These included dependent variables of self-efficacy and pattern language frequency as well as sophistication. In study two, reliability of observations scoring was also conducted between the researcher and volunteer observer (see section 3.16). In addition, reliability was augmented by comparisons made between the data sources as well including: the actual observation, the self-reporting from the PSTs of the lesson design and implementation, and the written lesson design documentation provided by the PSTs of the observed lesson.

Construct validity is concerned with whether the test items were indicators of the underlying construct in question. Underrepresentation is a danger where the test becomes too narrow. In the instance of explicit teaching or cooperative learning characteristics, the definitions were confirmed through literature reviews of explicit teaching and cooperative learning research based pedagogies used in the field.

3.18 Threats to External Validity

In study one, treatment interaction threats were considered as major threats to external validity. In study two these threats are considered minor as no treatment was delivered during study two.
Table 3.15

Threats to External Validity

<table>
<thead>
<tr>
<th></th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of representativeness of available populations / selection treatment interaction</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Interaction of setting and treatment</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Interaction of history and treatment</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

3.19 Procedures

3.19.1 Conditions.

The independent variable of embedded design and applied experience remained the same as for study one (see section 3.9.3). At the conclusion of study one data collection, an ethics application was prepared and submitted to the New South Wales State Education Research Applications Process (SERAP) for approval to visit Department of Education schools (see Appendix 9). SERAP approval was not granted until after the PST second year practicum had finished. The time difference between study one and two was 12 months.

3.19.3 Participants.

Volunteers for study 2 were recruited during a mandatory lecture for the PST practicum in semester 6 as part of a university program. PSTs were recruited by circulating a class list with columns for PSTs to indicate if they wished to participate in the study. Personal contact details were also sought at this time. The only information given to the PSTs about the research at this stage was that the researcher wished to observe a lesson using the pedagogies of explicit teaching or of cooperative learning. This ‘general’ information was delivered in a timeframe of 20 minutes across both campuses. The same protocols were followed across campuses with 20 volunteers from the embedded design cohort and 10 from the applied experience cohort volunteering for study. These 30 participants were a subset of the participants in study one. Table 3.16 provides a summary of the demographic information for participants in study two.
Table 3.16

Demographic Information of Participants in Study Two

<table>
<thead>
<tr>
<th></th>
<th>ED N (%)</th>
<th>AE N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 20-25</td>
<td>13 (63.00)</td>
<td>7 (72.72)</td>
</tr>
<tr>
<td>Age &gt; 26</td>
<td>7 (37.00)</td>
<td>3 (27.27)</td>
</tr>
<tr>
<td>Cohort</td>
<td>20 (77.98)</td>
<td>10 (22.02)</td>
</tr>
<tr>
<td>GPA</td>
<td>5.17</td>
<td>5.06</td>
</tr>
<tr>
<td>Average total points accrued in program</td>
<td>192</td>
<td>192</td>
</tr>
<tr>
<td>First degree of study</td>
<td>19 (94.11)</td>
<td>10 (100)</td>
</tr>
<tr>
<td>Some experience</td>
<td>5 (26.30)</td>
<td>4 (45.40)</td>
</tr>
<tr>
<td>Little experience</td>
<td>14 (68.40)</td>
<td>3 (27.00)</td>
</tr>
<tr>
<td>No experience</td>
<td>1 (5.80)</td>
<td>2 (25.00)</td>
</tr>
</tbody>
</table>

Note: ED and AE refer to the respective cohorts with ED being the experimental condition and AE being the control condition. An * is used to indicate statistical differences.

There were slightly younger PSTs in the applied experience condition with 73% (N=7) of participants under 25 years old; in the embedded design condition 63% (N=13) of participants fell into this age bracket. Their GPA scores remained close to each other and were 5.17 for the embedded design and 5.06 for the applied experience. The average total points accumulated in the program to date was exactly the same with embedded design and applied experience both being 192.

Participants were asked about their level of experience with people who have disabilities. With experiences ranging from teacher aide work or practicum experience to having a family member or friend with a disability: more than 95% (N=9) of participants in the embedded design condition had some experience, and 72% (N=7) of participants from the applied experience cohort had some experience of individuals with a disability. Many similarities are evident from this demographic information provided about the two groups. Compared to the larger sample found in study one, demographics such as first degree of study; campus of study; were similar. GPA scores were higher as more courses had been completed by all PSTs. In study two, there were less of the participants in the embedded design cohort who fell within in the under 25-years age group. Approximately 75% of the participants in the both cohorts again had some experience with others who had a disability.
3.19.3 Procedures for study two.

Once ethics approval for the study had been granted from SERAP (See Appendix 9), the study took place with a cohort of PSTs completing their third year practicum—which occurred in the sixth university semester of the program for a period of five weeks.

During the first meeting with volunteers for study two, procedures were elaborated to describe the observation and follow-up activities to be completed by the PSTs. A request was made by the researcher to have a copy of the written lesson plan prior to the lesson that would be analysed later. A short 15 to 20 minute narrative reflection of the lesson was also required from the PSTs following the delivery of the lesson.

Due to the time lapse between the end of study 1 and the beginning of study 2, the researcher determined it was necessary to provide an example of an explicit teaching lesson (for example: playing double bingo, see Appendix 8). Participants were required to write information from the demonstrated lesson under the different headings of characteristics associated with explicit teaching (anticipatory set, condition, action criteria, modelling, guided practice and independent practice; differentiation of process, content, product). The participants then completed the blank fidelity pro forma with this information following the demonstration (see Appendix 8). The completed information was not used any further.

The participants then needed to determine where on the standard lesson plan provided by the university they might include these details. Not all fields on the standard practicum form were relevant to the ET characteristics.

Next, a poor example of an ET lesson was demonstrated by the researcher where guided practice did not match the modelled section, for example. Volunteers were asked what was wrong with the demonstration lesson and how they might correct it if they had planned the lesson themselves. The feedback the PSTs provided here was verbal. The same discussion and examples were delivered to both groups of participants by the researcher.

The researcher organised practicum visits based on geographical locations of convenience in order to maximise the number of participants that might be involved.
Thirty PSTs were observed. The researcher called participants the night prior to each visit. PSTs were asked: “Is the lesson plan written and printed?” and “Are all the components of the lesson design in place?” No coaching was given; if questions were asked, the response was “Go back to the lesson designs of last session and check on desirable characteristics”.

3.20 S2 Data Collection Procedures

In study two, the data collection for both conditions involved one visit to the volunteer participants whilst they were on practicum. There were no further interventions for either the embedded design cohort or the applied experience condition in the intervening 12 months of university study.

The visit consisted of a reminder phone call the night prior to the lesson observation. The visit itself involved the PST providing the observer the lesson plan they had designed for the observation. The lesson was then observed. A reflection on their lesson was then sought considering their design and implementation of the same lesson. Following data collection, the observer provided feedback on both the design of the lesson plan and the lesson implementation for the PST.

The following Figure 3.5 illustrates the data collection sequence undertaken.

![Figure 3.5](image)

**Figure 3.5.** Data collection procedures for study two.

3.20.1 Dependent measures.

During study two, data were collected on self-efficacy and pattern language dependent variables and also dependent variables targeting the degree of design, implementation, and self-reflection fidelity. Study two dependent variables focused on the *instantiation of inclusive education tools in practice* as well as the maintenance of capacity building skills such as self-efficacy and the use of pattern language. The same instruments for self-efficacy and pattern language were used in study 2 as in study 1. An observation protocol was developed to measure PST design, implementation and self-reflection
fidelity of lessons delivered whilst on practicum. Each of these dependent measures will be elaborated in the following sections.

**3.20.1.1 Self-Efficacy questionnaire (SEIPD).**
SEIPD was used again in the second study for the purpose of establishing PST levels of self-efficacy. It was administered to the 30 volunteers who were a subset of PSTs who had been involved in the course for inclusive education the prior university year. The instrument was administered to volunteers during their initial practicum meeting, week 4 of semester 6. Reliability is reported in section 3.10.1.2 and validity for SEIPD is reported in section 3.10.1.3.

**3.20.1.2 Lesson Reflection used for pattern language**
Participants were asked to reflect on the design and also the delivery of the observed lesson that had been delivered whilst they were on their practicum placement. Participants were required to spend 15-20 minutes to write a reflection of their lesson design and implementation. No other instruction or guidance was given. This reflection was then analysed for pattern language by the researcher. The frequency (see section 3.11.2 of study 1) and sophistication (see section 3.11.3 of study 1) of the pattern language were able to be determined from the one reflection. Reliability for pattern language is reported in section 3.11.3.4 and validity is reported in section 3.11.3.5.

**3.20.1.3 Design Fidelity of the Lesson**
The researcher completed the design fidelity protocol based on lesson plans provided by the PST prior to lesson delivery. The same protocol was completed as a way of indicating whether explicit teaching characteristics were present in the design of the observed lesson (see Table 3.16).

**3.20.1.4 Implementation fidelity.**
Observation involves carefully planned and deliberate systematic examination of what is taking place, who is involved, and when and where things are occurring. Shaughnessy et al. (2009, p. 95) refers to a process of ‘scientific observation’ where the use of immediate awareness or direct cognition can yield more valid or authentic data than by other means (Cohen et al., 2007). It is a maxim in the social sciences that attitudes and behaviour are not always congruent so this is an important addition to the
data collection tools outlined above to answer the research questions noted earlier (Johnson & Christensen, 2008). It has the advantage over self-reports of the researcher’s ability to record actual behaviour rather than obtaining reports of intended behaviour. It also provides in-situ accounts rather than relying on second-hand information (Cohen et al., 2007).

Event recording was the technique used for measuring the occurrence of these specific behaviours and characteristics (Alberto & Troutman, 2012; Howell & Nolet, 2000; Johnson & Christensen, 2008; Zirpoli, 2015). This recording technique was appropriate as the behaviours or characteristics are defined as having a discrete beginning and end. Occurrence was represented on the likert scale as “absent … present” to note if a characteristic was evident in the lesson (i.e., key words, attention grabber). There was no need to count the frequency of occurrences, just to note if they were evident for the present/absent choice. A further scale of “incomplete … partially complete …. complete” (i.e., for elements of the learner outcome) was used. Here a score of ‘incomplete’ might be given if the condition, action or criterion was missing from the learner outcome. The score of ‘partially’ would be given if more than 50% of these elements were present; and a score of ‘complete’ given if 100% of the elements were present in a particular lesson. Scores of “infrequent … sometimes … frequent” were allocated for sections of guided practice in a similar manner. By allocating specific numerical amounts to the scale classifications, this allowed the data to be treated as ‘interval’ for each of the dimensions mentioned on the protocol (Shaughnessy et al., 2009). The items on the observation protocols for explicit teaching and also for cooperative learning were allocated a numerical score. For example, if the options were: ‘incomplete’, ‘partially complete’, and complete’, these were allocated a score of 0, 1 or 2 respectively. If items consisted of a scale of ‘absent’ or ‘present’ they were allocated a score of 0 or 1 respectively. The maximum score possible was 24 for design, implementation and self-reflection fidelity.

Observation was the data collection method, with information transferred to the observations protocol sheets following the lesson (Howell & Nolet, 2000; Zirpoli, 2015). Event recording was the scoring technique. Such specificity allows for defensible comparisons to be made across classrooms and between observers. The following table delineates the events for an explicit teaching lesson that were recorded.
Table 3.1
Characteristics of Explicit Teaching Research-Based Strategies

<table>
<thead>
<tr>
<th>Explicit teaching (ET)</th>
<th>Lesson purpose stated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anticipatory set created</td>
<td></td>
</tr>
<tr>
<td>Key words identified</td>
<td></td>
</tr>
<tr>
<td>Link to prior learning or background knowledge established</td>
<td></td>
</tr>
<tr>
<td>Goals /outcomes/objectives identified:</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td></td>
</tr>
<tr>
<td>Criteria</td>
<td></td>
</tr>
<tr>
<td>Teacher model provided</td>
<td></td>
</tr>
<tr>
<td>Guided practice /scaffolding</td>
<td></td>
</tr>
<tr>
<td>Independent practice</td>
<td></td>
</tr>
<tr>
<td>Differentiation of: condition, product, or content</td>
<td></td>
</tr>
</tbody>
</table>

Observations are also distinguished in the literature depending on whether they are ‘structured’ or ‘unstructured’ and ‘participant’ or ‘non-participant’ (Cohen et al., 2007). The current research involved non-participant structured observations as the protocol was carefully designed and organised into a checklist beforehand. Non-participant observations involved the researcher removing themselves as much as possible from the context under observation. Even though the actual presence would have an impact on behaviour observed, this was countered with triangulation of data from both the researcher and the PSTs.

3.20.1.5 Self-reflection fidelity
PST self-reflection was completed following delivery of the lesson. PSTs were asked to complete the same design fidelity protocol as a way of indicating their thoughts about the design of their lesson (see Appendix 8).

3.20.2.1 Reliability of the fidelity observations.
Zirpoli (2015) outlined several factors that could potentially affect the accuracy of data collected during direct observations: reactivity, observer drift, and recording procedure, location of the observations, observer expectancy, and characteristics of
those involved. Reactivity occurs when the participant is aware of being observed and changes in behaviour occur as a result. Thus, unobtrusive observations are best and the observer in this study spent time in the classrooms prior to the lesson being observed to minimise this effect on the PST as well as on the school students in the class.

Observer drift refers to a gradual shift in understanding and thus measurement of the target behaviours/characteristics. To control for this, a target behaviour needs to be defined very specifically, in terms that will remain clear throughout the program. Examples of the behaviour as well as non-examples of the behaviour were defined and used while reliability observations were being taken and compared. Cohen et al. (2007) make reference to a type of highly structured organisation called ‘pre ordinate observation’ where the characteristics have been worked out well in advance and the data will be used to confirm or refute hypotheses that have already been noted.

Recording was carried out in naturalistic settings of the classroom (Shaughnessy et al., 2009). Threats to reliability of observations were countered through the use of a trained observer taking duplicate observations and interrater reliability scores determined. The research assistant, acting as an observer, was trained regarding the definition of the target behaviour, and observations were made as unobtrusively as possible although it was not possible to make frequent and systematic observations of the same PST. With any scoring procedure, it is imperative to practice completing the interval schedule of observations (Cohen et al., 2007). High reliability scores were an indication that the target behaviours (characteristics) were clearly defined. Agreements were calculated using the formula of agreements divided by agreements plus disagreements to yield a percent agreement statistic.

3.20.2.2 Validity of the fidelity observations.
Cohen et al. (2007) refer to observations as having ‘ecological validity’ (p. 396) as they are taken in natural settings and allow the context of programs to be understood. The focus of this study was not on potential data gathering of: the physical setting, the human setting, or the interactional setting; but rather on the program setting where resources and their organisation and pedagogical styles may be observed. The observation protocol is seen to have ‘content validity’ as it faithfully represents the objectives of an instructional sequence (Burns, 2000, p. 351). It is concerned with how well the items (content) on the observation protocol represented research-based
characteristics of inclusive education pedagogies; and how well the items measured PST observable mastery of these concepts. While it did not claim to represent the ‘universe’ of the content being measured, it did represent the universe of the content set out in the learning objectives for the course. The content validity of the implementation protocol was enhanced as the observation represents the actual content that was taught across the weeks of the university teaching session. In the inclusive education course this was evidenced as the reading objectives given to PSTs each week which were also the focus of lecture and tutorial activities. These same objectives were framed as questions used in the quiz assessments. This repetition was part of the embedded design taken from the theoretical driver. In this way, the content validity is representative of the sampling adequacy and is an extension of the content covered in classes.

The observation protocol might be argued to have high construct validity, which refers to how well the test measures PSTs’ ability to implement the pedagogies used in inclusive education. Research-based pedagogies were noted in the literature as being effective for use within inclusive classrooms with students who have a wide variety of learning needs. The strategies included: explicit teaching, cognitive strategies, cooperative learning, and differentiation of these in three ways (process, product and content).

The observations were seen to have high face validity. Even though face validity may be desired it is not as important to generalisability of results as the other forms of validity. It is also difficult, if not impossible, to measure a validity of this type. To enhance face validity in this research, the volunteers were taken through introductory activities at each campus prior to visits so the observation protocol could be explained and linked to their current practicum lesson plan pro forma required by the university.

3.21 Data Analysis

3.21.1 Self-efficacy analysis.
Study two relied on matched samples of participants for their SEIPD descriptive statistics (means and standard deviations). In study 2, it was not possible to use inferential statistics due to the small number of participants involved (ED = 20, AE = 10).
3.21.2 Pattern language frequency and sophistication.

Reflections from study 2 were analysed in the same way as study 1 to establish pattern language frequency (see section 3.11.2) and sophistication (see section 3.11.3). The frequency count was analysed from the reflection the PSTs wrote following delivery of a lesson whilst they were on practicum. No inferential statistical analysis was possible for study 2 as the number of participants was too low. Instead, the descriptive statistics from matched samples of participants (from study 1 and 2) were compared. For example the descriptive statistics from the matched participants in study 1 were compared to the descriptive statistics for the same participants in study 2. The PL content area for comparison of frequency and sophistication was that of explicit teaching as this was the only form of lesson that was observed in study 2. That is, lessons involving ‘cognitive strategies’ or ‘cooperative learning’ were not observed by the researcher.

3.21.3 Design fidelity for explicit teaching.

Reflections used in written documentation are also referred to as artefacts (Cohen et al., 2007). In this thesis they refer to lesson plans developed by the PSTs. Collection of these artefacts added important contextual details to the data available for analysis. No set lesson plan design was insisted upon in the conduct of this research as some PSTs were required to write full lesson plans based on the outline provided by the university practicum office, and some were simply required to write day book notes. The only stipulation from the researcher was that the observed lesson plan designs have the headings of required characteristics for the pedagogy being used.

Document analysis of the lesson designs was carried out using the fidelity protocol as the real time lesson implementation for triangulation of the findings. That is, the design of lessons delivered by PSTs was scored by the observer and analysed using the same protocol as the implementation fidelity protocol (see 2.21.4).


The researcher used the observation protocol developed for the purpose of this research to observe a 30- to 40-minute class session. The PST invited the observer to attend the school setting at a particular time to view the lesson. Unlike the regular visits by university mentors, these observations were undertaken specifically for the
purposes of this study. Only one visit occurred for each participant throughout the practicum experience. These observations occurred following the midpoint of ten days when the university designated mentor had already visited and discussed satisfactory progression of the PST with the classroom teacher. The visits for this study had no impact on the grade the PST was to receive for their practicum course.

The observation protocols included items that described characteristics of the practice under observation. For example, the explicit teaching protocol (11 core items plus 3 more for differentiation, see Appendix 6) required the observer to determine whether or not the lesson purpose was stated, the anticipatory set created and key words identified. Links to prior learning had to be established, then outcomes/objectives identified. Was a clear teacher model evident with logical steps? Did guided practice follow that was also broken down into logical steps and related to the model? Did all students complete independent practice? Was there evidence of differentiation of condition, product, or content? The observer recorded the presence or absence along a continuum for the essential characteristics of the practice by designating them as: complete … partially complete … incomplete. These responses were allocated scores of complete = 2; partially complete = 1 and incomplete = zero. The protocols were designed to report the integrity with which the instructional design was implemented. The raw scores for sections were analysed as well as total raw scores.

Implementation integrity for each observation was reported based upon the items observed as present along the Likert scales of the observation protocol. The scales did not attempt to establish psychometric validity as they used criterion referenced standards from the research literature. Given the data from this tool categorical in nature, Chi-square analysis using contingency table was employed (see Appendix 6).

3.21.5 Self-reflection fidelity for explicit teaching.
Following lesson delivery, the PST was required to fill in the same scoring protocol as a means of self-reflection of their lesson design and implementation fidelity that had been used by the researcher. This was then used as a focus for discussion between researcher and PST. This discussion was not used as part of the data analysis.
This chapter presented the research design and processes employed across study one and study two; provided descriptions of the independent and dependent variables applicable to both studies; threats to validity have been considered in the case of each dependent variable, as well as a discussion of means to minimise these threats. Data collection tools and procedures have been explained and the analysis procedures of the same are included. The following chapter presents the results for both studies.
Chapter 4: Results

This chapter presents the results for both studies which examined the effectiveness of a theoretically driven course design. Comparisons were made between a course based on the *embedded design* principle and another based on an *applied experience* design. Research-based capacity building skills and attitudes were investigated in study one and then the instantiation of these characteristics in a practicum setting were investigated for study two. Both descriptive and inferential statistics are presented to address the primary research questions. First, for study one, group equivalence at pre-test is described. Second, the effects of each intervention condition (*embedded design* and *applied experience* design) on mastery of course content, self-efficacy, and sophistication of pattern language were examined over time using repeated measures analyses of variance (ANOVA). These effects were determined using the dependent variables of: pedagogical content knowledge (PCK), frequency and sophistication of pattern language used, and self-efficacy. Finally, the effect of each condition on design of lesson plans, implementation integrity of those lessons and a self-reflection from the PSTs about the lesson was examined in study two. Results for study one will be presented first followed by results for study two.

4.1 Study One: Group Equivalence at Pre-test

Prior to the commencement of intervention, participants were assigned to one of two conditions: *embedded design* or *applied experience* design. Participants were assigned to intervention conditions by intervention cohort and were grouped based on convenience. To detect pre-intervention differences across conditions for each of the demographic variables used in the study and for each of the dependent variables, two analyses were conducted: two-way contingency table analysis and independent samples t-tests.

Two-way contingency table analyses were conducted to detect pre-intervention differences between conditions on demographic variables including: age; previous experiences with individuals who have a disability; and degree of teacher training completed ($\alpha = .05$). Prior to running the analyses, the following assumptions were examined: (i) independence between variables; and (ii) homogeneity of proportions. Both assumptions were met for each of the demographic variables with the exception
of assumption two for degree of study. For this variable, the sample size was fewer than five cases in two cells.

Results of the two-way contingency table analysis revealed no statistically significant difference between the two intervention conditions—*embedded design* and *applied experience* design—for degree of study completed \( \chi^2 (1, N = 134) = 1.09, p = .30 \), or previous experience with individuals who have a disability \( \chi^2 (2, N = 134) = 3.96, p = .14 \).

A statistically significant difference was, however, detected between age groups across intervention conditions, \( \chi^2 (2, N = 134) = 14.29, p = .001 \). Of the participants in the study, 79% of pre-service teachers (PSTs) from the *embedded design* condition fell within the 19-29 age group; 12% within the 30-39 age group; and 9% within the 40-49 age group. For the *applied experience* condition, 49% of PSTs fell within the 19-29 age group; 17% within the 30-39 age group; and 34% within the 40-49 age group. Differences on this demographic variable were not perceived to be of major concern as there were no statistically significant differences between PSTs across intervention conditions on all pre-test variables as described below.

Independent-samples *t*-tests were conducted to detect pre-intervention differences across intervention conditions in demographic variables including: (i) grade point average (GPA); and (ii) total points accrued in the teacher education program. In addition, *t*-tests were also conducted to detect differences in pre-test performance on all dependent measures including: SEIPD; pedagogical content knowledge; and pattern language frequency. Sophistication of pattern language was not measured at pre-test. Prior to conducting each *t*-test the assumptions of this analysis were investigated including: (i) the test variable is normally distributed in each of the two populations; (ii) the variances of the normally distributed test variable for the populations are equal; and (iii) the cases represent a random sample from the population. Results indicated that the test variables were normally distributed in each population for GPA scores, SEIPD, pedagogical content knowledge, and pattern language frequency. While the sample sizes were unequal, for the most part they did conform to normality for the test variables. However, the assumption was violated for ‘total points accrued’. Total points accrued reflect the number of courses a PST had undertaken in their program of study, therefore the stage in their progress towards becoming a teacher. This simply
indicated that some PSTs were studying the inclusive education course at differing time points within their degree. The second assumption was assessed by examining Levene’s test for equality of variances and data was reported using the \( t \) value that assumes unequal variance where required. The third assumption, that the cases represent a random sample from the population, was met as the population that participated in this research were a random sample from within the university population.

Results of the independent samples \( t \)-tests revealed no statistically significant differences between the embedded design or applied experience design conditions at pre-test across: GPA \( t(132) = .48, p = .63 \), total course points accrued in the program \( t(132) = .1.52, p = .13 \); pedagogical content knowledge (PCK) \( t(132) = -.30, p = .76 \); and pattern language frequency \( t(130) = -1.47, p = .15 \).

The effect of each intervention condition on the dependent variables will now be discussed. First, discussion focuses on study one concerning variables of: pedagogical content knowledge, use of pattern language and self-efficacy; and second, discussion focuses on the design and translation of these skills and attitudes into classroom practice in study two.

4.1.1 Reliability of instructor implementation

The following table (Table 4.1) represents the degree of agreement between the three instructors teaching the inclusive education course. Complete agreement was achieved for weeks: 1, 2, 3, 5, 7, 8, 9 and 11. No lecture was held weeks 4, 6, 10, and 12. Agreement was somewhat less consistent for the tutorials with the range in agreement being from 75% to 100% with 9 of the 12 tutorials above the 80% threshold.

Some variability existed in the tutorials however this was not considered critical as the differences occurred between a strong level of agreement and a medium level of agreement. The differences did not occur for critical course content, but rather the emphasis with which the instructor covered strategies such as: discussing a grading rubric component relevant to the content of the week, or providing an example of the pedagogy being explored in a particular week.
Table 4.1
Reliability of instructor implementation

<table>
<thead>
<tr>
<th>Lectures</th>
<th>Agreements</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5/5</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>9/9</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>5/5</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>No lecture</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>8/8</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>No lecture</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>7/7</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>5/5</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>6/6</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>No lecture</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>7/7</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>No lecture</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tutorials</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5/6</td>
<td>83</td>
</tr>
<tr>
<td>2</td>
<td>5/6</td>
<td>83</td>
</tr>
<tr>
<td>3</td>
<td>3/4</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>2/2</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>3/4</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>6/6</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>3/3</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>4/4</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>3/4</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>1/1</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>2/2</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>4/4</td>
<td>100</td>
</tr>
</tbody>
</table>

4.2 Application of the Embedded Design Principle to the Design of a Pre-service Teacher Education Inclusive Education Course

The research questions from study one sought to examine the effect of each intervention condition (embedded design and applied experience) over time on student mastery of course content, improvement in self-efficacy and sophistication of pattern language used. The research hypothesis for study one were as follows:

H₁ = There will be a statistically significant difference in the levels of:

a) The mastery of course content in the form of pedagogical content knowledge (PCK)

b) The frequency and sophistication of pattern language, and the level of

c) Self-efficacy between the PSTs who experienced application of the embedded design principle to the theoretical design of an inclusive
education course as opposed to those who participate in an *applied experience* condition.

\[ H_{01} = \] There will be no statistically significant difference in the levels of:

a) The mastery of course content in the form of pedagogical content knowledge (PCK)

b) The frequency and sophistication of pattern language, and the level of

c) Self-efficacy between the PSTs who experience the application of *embedded design* principle to the theoretical design of the inclusive education course as opposed to those who participate in an *applied experience* condition.

A series of two-way repeated measures ANOVA were conducted for each set of measures with intervention condition (*embedded design* versus *applied experience*) as a between subjects factor and time (pre and post, or weeks 4, 8 and 12) as a within subject factor. The analyses were conducted using SPSS software (SPSS Inc, 2002). When conducting repeated measures ANOVA, SPSS software provides output for the standard univariate test, alternative univariate tests (degrees of freedom corrected) and a multivariate test. As output from the standard univariate test could not have been used in this instance (the assumption of sphericity was not met), the output from the multivariate tests was used. In addition, this output was chosen because multiple repeated measures ANOVA were conducted, thereby increasing the risk of a Type 1 error. The multivariate test provided a more conservative estimate of significance and was used to correct for possible error (Green & Salkind, 2008).

One-way repeated measures ANOVA and paired samples *t*-tests were conducted where appropriate as a follow up to statistically significant interactions and main effects. Holm’s sequential Bonferroni procedure was used to control for Type 1 error across the pairwise comparisons (\( \alpha = .05 \)) for the paired samples *t*-tests. When applying Holm’s sequential Bonferroni correction for up to three pairwise comparisons, alpha was adjusted to .0166 for the initial statistically significant comparison (\( \alpha = .05/3 \)), .0025 for the second statistically significant pairwise comparison (\( \alpha = .05/2 \)), and .05 for the final statistically significant pairwise comparison (\( \alpha = .05/1 \)).
Prior to commencing analysis of the data, assumptions of repeated measures ANOVA were tested. First, the multivariate assumption that the difference scores for each dependent measure were multivariately normally distributed in the population was tested. This assumption assessed the value of the dependent variable at each level of the within-subjects and each level of the between-subjects variable and its normal distribution in the population. This assumption was satisfied based on visual inspection of histograms and P-plot graphs for each of the dependent measures. Repeated measures ANOVA is considered robust to violations of the normality assumption if the sample size is adequate. For these analyses, the sample size was adequate; however, the group sizes were not balanced (ED, N=85 and AE, N=24). Outliers in the data set were not considered to be severe. One condition had one-third the population of the other so the six cells involved in this 3x2 design were unlikely to have an even distribution (Green & Salkind, 2008, p. 247). The skewness and kurtosis statistics for all distributions were close to zero. Z_s and Z_k were within +/- 1.96 for all groups suggesting all groups were approximately normal (Allen & Bennet, 2010, pp. 80, 113). The Shapiro-Wilk statistic (W) ranged from .93-.97 so it was concluded that the assumption of normality was not violated for any of the groups of scores, even though the sample sizes were uneven.

The second assumption of repeated measures ANOVA is that individuals represent a random sample from the population. The participants from the study were not randomly sampled from within the broader university population. Assignment to intervention condition was based on convenience sampling technique. Data analysis conducted to determine group equivalence at pre-test suggested intervention groups were equivalent on all but the ‘age group’ pre-test variable, even though random assignment to condition was not possible. The second assumption also requires that difference scores for any one individual be independent from the scores for any other individual. This assumption was met in all instances.

Effect sizes were calculated to examine the difference between the mean scores for each of the dependent measures. For repeated measures ANOVA, effect sizes were reported as eta squared (\(\eta^2\)), an index of the proportion of variance explained by a variable (Grimm & Yarnold, 1995). Effect sizes for the multivariate analysis were reported as the multivariate eta square using Wilks’s lambda as the statistic for a main or interaction effect. Means of the factors ranged from 0 to 1, with .01 as small, .09 as
medium, and > .25 as large (Green & Salkind, 2008). Effect size analyses were reported for each statistically significant interaction, simple effect, and main effect.

4.3 Mastery of Course Content in the Form of Pedagogical Content Knowledge

Improvement in pedagogical content knowledge was assessed at pre- and post-test for each condition using knowledge test data. Table 4.1 provides descriptive statistics for PCK data for each intervention condition. Data in Table 4.1 reveal similarity in student PCK at pre-test and the differences between groups at post-test in favour of the embedded design condition.

Table 4.2
Pedagogical Content Knowledge Pre-Post Mean and Standard Deviation Scores by Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded design</td>
<td>84</td>
<td>3.26</td>
<td>26.22</td>
</tr>
<tr>
<td>Applied experience</td>
<td>22</td>
<td>3.45</td>
<td>18.82</td>
</tr>
</tbody>
</table>

N.B. 1 participant from the embedded design condition and 2 from the applied experience condition did not complete both pre and post-tests.

A one-way repeated measures ANOVA was conducted with intervention condition (embedded design versus applied experience) as a between subject’s factor and time (pre-test versus post-test) as a within subjects factor (Allen & Bennett, 2010, p. 114). The Condition x Time interaction was examined using the multivariate criterion of Wilks’s Lambda (Λ). T-tests were used as a simple effects analysis to follow on from a statistically significant interaction. Holm’s sequential Bonferroni procedure was used to control for Type 1 error across comparisons for simple effects with alpha set at .01.

Pedagogical content knowledge at pre- and post-test showed the Condition x Time interaction was statistically significant, Λ = .73, F(1, 104) = 38.17, p < .01, multivariate η² = .27. Follow up analysis for simple effects were conducted using paired-samples t-tests. Significant differences were detected between pre and post for both intervention conditions: embedded design t(83) = -44.10, p < .01; and applied experience t(21) = -11.31, p < .01. Independent-samples t-tests were conducted to follow up the statistically significant interaction between groups (intervention
conditions). Hedges’ g is used as opposed to Cohen’s d as it provides a measure of effect size that is weighted according to the relative size of each sample and is an alternative where there are different sample sizes being compared. Results from this analysis revealed no statistical differences between intervention conditions at pre-test, \( t(109) = -0.57, p = .57 \) and statistically significant differences between intervention conditions at post-test, \( t(104) = 5.89, p < .01, g = 1.41 \). This standardised measure of effect size refers to 0.2 as being ‘small’ in magnitude, measures around 0.5 as being ‘medium’, and those around or above 0.8 as being ‘large’ in magnitude.

Figure 4.1 illustrates student performance at pre- and post-test by intervention condition. Data in both Table 4.1 and Figure 4.1 reveal similarity in PST pedagogical content knowledge at pre-test. For study one, both the table and the figure highlight the differences between groups at post-test in favour of the embedded design condition as revealed in the statistical analysis.

![Figure 4.1. Pedagogical content knowledge at pre- and post-test occasions by condition for study one.](image-url)
Improvement in pedagogical content knowledge was also assessed during the teaching semester of the intervention in Weeks 4, 8 and 12 for each condition using mastery quiz data. For both conditions there was an expectation of mastery (set at 80%) of the content in order to satisfy passing criteria of the quiz assessment item. Table 4.2 provides descriptive statistics for pedagogical content knowledge data for each intervention condition at each quiz occasion.

Table 4.3

<table>
<thead>
<tr>
<th>Quiz Occasion Pedagogical Content Knowledge Mean and Standard Deviation Scores by Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Embedded design</td>
</tr>
<tr>
<td>Applied experience</td>
</tr>
</tbody>
</table>

N.B. 2 participants from the applied experience condition did not complete all three quizzes.

A two-way repeated measures ANOVA was conducted with intervention condition (embedded design versus applied experience) as a between subjects factor and time (week 4 versus week 8 versus week 12) as a within subjects factor. The Condition x Time interaction was examined using the multivariate criterion of Wilks’s Lambda (Λ). A one-way repeated measures ANOVA was used to follow from a statistically significant interaction. Holm’s sequential Bonferroni procedure was used to control for Type 1 error across comparisons for simple effects that followed from the one-way ANOVA with alpha set at a conservative .01 level.

Pedagogical content knowledge was assessed using the mastery quizzes at weeks 4, 8 and 12. The Condition x Time interaction was statistically significant, Λ = .58, \( F(2, 104) = 37.84 \), \( p < .01 \), multivariate \( \eta^2 = .42 \).

A one-way repeated measures ANOVA was conducted to follow up the statistically significant interaction. Results from this analysis revealed statistically significant differences across mastery quizzes for the entire cohort, Λ = .59, \( F(2, 105) = 36.06 \),
Follow up to a significant omnibus ANOVA, pairwise comparisons were reported. Specifically, the pairwise comparisons revealed no significant difference between quiz 1 ($M = 10.58, SD = 1.60$) and 2 ($M = 10.36, SD = 1.40$). Significance was evident between quizzes 1 and 3 ($M = 8.38, SD = 3.14$); and also quizzes 2 and 3. Paired samples $t$-tests confirmed no significance between quizzes 1 and 2, $t(108) = 1.46, p = .15$; significance between quizzes 2 and 3, $t(106) = 8.10, p < .05$; and statistically significant difference between quizzes 1 and 3 $t(106) = 8.25, p < .05$.

A one-way repeated measures ANOVA was used to explore differences between intervention condition scores on the three quizzes. For the *embedded design* condition, $\Lambda = .69, F(2, 83) = 17.93, p < .01$, multivariate $\eta^2 = .30$. Results of pairwise comparisons for the *embedded design* condition mirrored the entire cohort, with no significant difference between quizzes 1 and 2, $t(84) = 1.09, p = .28$. There were significant differences between quizzes 1 and 3, $t(84) = 5.67, p < .05$; and quizzes 2 and 3, $t(84) = 5.47, p < .05$.

One way repeated measures ANOVA for the *applied experience* condition revealed significant differences between quiz occasions, $\Lambda = .12, F(2, 20) = 73.59, p < .01$, multivariate $\eta^2 = .88$. Results of pairwise comparisons for the *applied experience* condition mirrored the entire cohort, with no significant difference between quizzes 1 and 2, $t(23) = 1.03, p = .31$. There were significant differences between quizzes 1 and 3, $t(21) = 11.52, p < .05$; and quizzes 2 and 3, $t(21) = 11.47, p < .05$.

Independent samples $t$-tests for each quiz were conducted. Occasion showed significant difference in favour of the *embedded design* condition. Quiz 1/week 4, $t(109) = 3.67, p < .01$; quiz 2/week 8, $t(107) = 4.90, p < .01$; and quiz 3/week 12, $t(105) = 9.82, p < .01$, $g = 2.33$. Hedges’ $g$ was used to provide a measure of effect size at quiz 3 occasion as it is weighted according to the relative size of each sample.

Figure 4.2 illustrates student performance at each quiz occasion by intervention condition. Data in both Table 4.2 and Figure 4.2 highlight the differences between groups at each quiz occasion in favour of the *embedded design* condition as revealed in the statistical analysis. The results for the *embedded design* condition remained relatively stable across quiz occasions with a slight decrease at quiz 3. The results for
the *applied experience* condition began at a lower mean but were stable across quizzes 1 and 2. This condition then decreased significantly at quiz 3 occasion. There was a mastery requirement of the course for the quizzes which means that higher scores were expected under both conditions.

These results indicate that hypothesis H1a was supported as there were statistically significant differences between intervention conditions in favour of *embedded design* at post-test for mastery of pedagogical content knowledge. The hypothesis was also supported with results from the quizzes with significant differences at each quiz occasion in favour of the *embedded design* condition.

![Graph showing means of quiz scores by condition for study one.](image)

*Figure 4.2. Quiz occasion pedagogical content knowledge mean scores by condition for study one.*

### 4.4 Frequency and Sophistication of Pattern Language

The frequency and sophistication of pattern language over time as a result of each intervention condition was assessed using two measures: (i) pattern language frequency on pedagogical content knowledge assessments pre- and post-intervention
and quiz reflections at weeks 4, 8 and 12 of the semester; and (ii) solo level analysis of PST quiz reflections at weeks 4, 8 and 12 of the semester. SOLO level of analysis only involved data collection at each of the three quiz occasions and did not involve a pre and post-test as the frequency of PL.

4.4.1 Frequency of pattern language at pre-post occasion.

Improvement in pattern language frequency was assessed at pre- and post-test for each condition using knowledge assessment data. The same student narrative/answers that were analysed for pedagogical content knowledge at pre- and post-test were used to analyse the use of pattern language frequency and sophistication. Table 4.3 provides descriptive statistics for pattern language frequency data for each intervention condition. The results for PL frequency will be reported for study one initially, and then for study two.

Table 4.4

Pattern language Pre-Post Mean and Standard Deviation Scores by Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Embedded design</td>
<td>83</td>
<td>1.28</td>
</tr>
<tr>
<td>Applied experience</td>
<td>22</td>
<td>1.64</td>
</tr>
</tbody>
</table>

N.B. 2 participants from the embedded design condition and 3 from the applied experience condition did not complete both pre and post-tests.

A two-way repeated measures ANOVA was conducted with intervention condition (embedded design versus applied experience) as a between subjects factor and time (pre-test versus post-test) as a within subjects factor. The Condition x Time interaction was examined using the multivariate criterion of Wilks’s Lambda (Λ). T-tests were used as a simple effects analysis to follow on from a statistically significant interaction. Holm’s sequential Bonferroni procedure was used to control for Type 1 error across comparisons for simple effects with alpha set at the .01 level.

For frequency of pattern language use at pre- and post-test, the Condition x Time interaction was statistically significant, Λ = .81, F(1, 103) = 24.82, p < .01, multivariate η² = .19. Follow up paired-samples t-tests for simple effects revealed statistically significant differences between pre- and post-test for both intervention
conditions. *Embedded design* resulted in $t(82) = -40.50, p < .001$; and the *applied experience* resulted in $t(21) = -12.55, p < .001$.

Independent-samples $t$-tests were conducted to follow up the statistically significant interaction for between groups (intervention conditions). Results from this analysis revealed no significant differences between intervention conditions at pre-test, $t(107) = -1.39, p = .19$ and statistically significant differences between intervention conditions at post-test, $t(105) = 4.84, p < .001, g = 1.16$.

Figure 4.3 illustrates student performance at pre- and post-test by intervention condition. Data in both Table 4.3 and Figure 4.3 reveal similarity in student use of pattern language at pre-test and highlight the differences between groups at post-test in favour of the *embedded design* condition as revealed in the statistical analysis. These results support the hypothesis $H_{1b}$ as the frequency of pattern language use at post-test occasion was significantly higher for the *embedded design* cohort.

*Figure 4.3. Pattern language frequency pre-post mean scores by condition.*
4.4.2 Frequency of pattern language at quiz occasion.

Improvement in pattern language frequency was also assessed at quiz occasion for each condition using reflections written in conjunction with quizzes from weeks 4, 8 and 12. Table 4.4 provides descriptive statistics for pattern language frequency data for each intervention condition across quiz occasions.

Table 4.5

| Quiz Occasion Pattern Language Mean and Standard Deviation Scores by Condition |
|-------------------------------|-----------------|-----------------|-----------------|
| Condition                     | Week 4 | SD  | Week 8 | SD  | Week 12 | SD |
| Embedded design               | N=80   | 4.08| 2.72   | 9.27| 3.85   | 10.44 | 6.46 |
| Applied experience            | N=23   | 2.43| 1.70   | 6.09| 3.25   | 5.22  | 3.64 |

N.B. 5 participants from the embedded design condition and 1 from the applied experience condition did not complete all quizzes.

A two-way repeated measures ANOVA was conducted with intervention condition (embedded design versus applied experience) as a between subjects factor, and time (week 4, week 8 and week 12) as a within subjects factor. The Condition x Time interaction was examined using the multivariate criterion of Wilks’s Lambda (Λ). A one-way repeated measures ANOVA was used to follow up from a statistically significant interaction. Independent samples t-tests were used for simple effects analysis to follow on from a statistically significant result across time for each condition with Holm’s sequential Bonferroni procedure used to control for Type 1 error. Alpha was adjusted to .01 for each analysis.

Results for pattern language frequency scores indicate the Condition x Time interaction was statistically significant, Λ = .91, F(2,100) = 5.08, p = .008, multivariate η² = .09. A one-way repeated measures ANOVA was conducted to detect differences in pattern language frequency across occasion for the entire cohort. Results indicate statistically significant differences in pattern language frequency across occasion, Λ = .30, F(2,101) = 117.59, p < .05, multivariate η² = .7. Pairwise comparisons from the statistically significant ANOVA revealed there was a -4.86 difference between means for pattern language frequency between quiz 1 (explicit teaching) and quiz 2 (cognitive...
strategies) which was statistically significant $t(105) = -14.45, p < .05$; a -5.56 difference between means for pattern language frequency between quiz 1 (explicit teaching) and quiz 3 (cooperative learning) with statistical significance $t(104) = -11.07, p < .05$; and a -7.71 difference in means between pattern language frequency in quiz 2 (cognitive strategies) and quiz 3 (cooperative learning) which was not significant $t(104) = -1.18, p = .24$.

Follow up simple effects analysis were conducted using one-way repeated measures ANOVA to detect differences in pattern language frequency across occasion for the embedded design condition. Results indicate statistically significant differences in pattern language frequency for the embedded design condition over time, $\Lambda = .26, F(2,78) = 108.44, p < .05$. As a follow up to the significant one-way repeated measures ANOVA, paired samples $t$-tests for the embedded design cohort revealed statistically significant differences between pattern language frequency in quiz 1 and quiz 2, $t(81) = -13.60, p < .05$; significant differences between pattern language frequency between quiz 1 and quiz 3, $t(81) = -10.77, p < .05$; and no significant differences between pattern language frequency in quiz 2 and quiz 3, $t(81) = -1.67, p = .09$.

Follow up simple effects analysis was conducted using one-way repeated measures ANOVA for the applied experience condition. This revealed a statistically significant difference in pattern language frequency over time, $\Lambda = .38, F(2,21) = 16.56, p < .05$. As a follow up to the significant one-way repeated measures ANOVA, paired samples $t$-tests for the applied experience cohort revealed statistically significant differences between pattern language frequency in quiz 1 and quiz 2, $t(23) = -5.63, p < .05$; statistically significant differences in pattern language frequency between quiz 1 and quiz 3, $t(22) = -4.13, p < .05$; and no statistically significant differences between pattern language frequency use in quiz 2 and quiz 3, $t(22) = 1.19, p = .24$.

To detect differences between intervention conditions for pattern language frequency at each quiz occasion, independent-samples $t$-tests were conducted to follow up the statistically significant interaction. Results from this analysis revealed statistically significant differences between intervention conditions at quiz 1, $t(61.12) = 3.49, p < .05$ in favour of the embedded design condition. There was significant difference at quiz 2, $t(106) = 3.92, p < .05$ in favour of the embedded design; and also at quiz 3, $t(64.39) = 4.67, p < .05, g = 0.87$ in favour of the embedded design condition. Note
that Levene’s Test for Equality of Variances were violated for quiz 1 and quiz 3 samples so results for ‘equal variances not assumed’ were reported. These quiz results support hypothesis H$_{1a}$ as results for the embedded design cohort were significantly higher.

Figure 4.5 illustrates student performance at quiz occasion by intervention condition. Data in both Table 4.4 and Figure 4.5 highlight the differences between groups at post-test in favour of the embedded design condition as revealed in the statistical analysis. These results also support the hypothesis H$_{1a}$. The mean scores increased across quiz occasions for the embedded design condition. They also increased for the applied experience condition across quizzes 1 and 2 but halved for quiz 3. The content covered each time was different and suggests the difficulty of the content increased each quiz for PST educators in both intervention conditions.

![Figure 4.4](image)

*Figure 4.4. Mean scores for pattern language by condition.*

4.4.3 Sophistication of pattern language at quiz occasion.

The sophistication of the pattern language was assessed at quiz occasion for each condition using reflections written in weeks 4, 8 and 12 to determine if the
sophistication of pattern language improved differentially with each teaching cycle implemented. Table 4.6 provides descriptive statistics for pattern language SOLO data for each intervention condition across quiz occasions.

Table 4.6

<table>
<thead>
<tr>
<th>Quiz Occasion</th>
<th>Pattern language Sophistication Mean and Standard Deviation Scores by Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Occasion 1</td>
</tr>
<tr>
<td></td>
<td>Week 4</td>
</tr>
<tr>
<td>Condition</td>
<td>N</td>
</tr>
<tr>
<td>Embedded design</td>
<td>80</td>
</tr>
<tr>
<td>Applied experience</td>
<td>23</td>
</tr>
</tbody>
</table>

N.B. 5 participants from the embedded design condition and 1 from the applied experience condition did not complete all quizzes.

A two-way repeated measures ANOVA was conducted with intervention condition (embedded design versus applied experience) as a between subjects factor and time (week 4, week 8 and week 12) as a within subjects factor. The Main effect for Time was examined using the multivariate criterion of Wilks’s Lambda (Λ). Pairwise comparisons were used as a simple effects analysis to follow on from a statistically significant result across time for each condition with Holm’s sequential Bonferroni procedure used to control for Type 1 error. Independent sample t-tests were used to determine differences between conditions. Alpha was adjusted to .01 for each analysis.

Analysing the SOLO level raw scores on quiz results at weeks 4, 8 and 12 of the semester, the Condition x Time interaction was not statistically significant, Λ = .99, $F(2,99) = .06, p = .94$, multivariate $\eta^2 = .001$. There was a statistically significant main effect for Time, $\Lambda = .62, F(2,99) = 30.23, p < .05$, multivariate $\eta^2 = .38$. Follow up pairwise comparisons on the main effect for time revealed there was a -.79 difference between means for SOLO levels in quiz 1 (explicit teaching) and quiz 2 (cognitive strategies) which was statistically significant $t(104) = -5.75, p < .05$; a 1.10 difference in means between pattern language SOLO levels in quiz 2 (cognitive strategies) and quiz 3 (cooperative learning) with statistical significance $t(103) = 9.23, p < .05$; and a .31 difference between means for SOLO levels in quiz 1 (explicit teaching) and quiz 3 (cooperative learning) which was not statistically significant, $t(103) = 1.84, p = .07$. 
The independent samples $t$-test analysis revealed no statistical differences were noted between the intervention conditions for sophistication of pattern language being used at each quiz occasion. Quiz 1: $t(107) = .47, p = .63$; quiz 2: $t(105) = 1.55, p = .12$; and quiz 3: $t(104) = .89, p = .37, g = 0.23$.

Figure 4.6 illustrates student performance at quiz occasion by intervention condition. Data in both Table 4.6 and Figure 4.6 reveal no significant interaction between quiz reflections in terms of the sophistication of pattern language being used by PSTs. Both conditions improved in the level of sophistication of pattern language being used from quiz 1 to quiz 2 and both declined sharply in quiz 3, reflected by the increased difficulty of the final quiz. These findings support the null hypothesis $H_{01b}$ as there were no significant differences between conditions for the sophistication of pattern language being used in the quizzes.

![Figure 4.5](image_url)  
*Figure 4.5. Quiz occasion pattern language sophistication mean scores by condition.*
In summary, the hypothesis $H_{1b}$ was supported for pattern language frequency used by PSTs, but the null hypothesis $H_{01b}$ was supported for the sophistication of that pattern language being used by PSTs.

4.5 Self-efficacy

Improvements in self-efficacy were assessed at pre- and post-test for each condition using the SEIPD scale (Self-Efficacy toward Interactions with People with Disabilities). Table 4.6 provides descriptive statistics for self-efficacy data for each intervention condition at each occasion.

<table>
<thead>
<tr>
<th>Table 4.7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-efficacy (SEIPD) Mean and Standard Deviation Scores by Condition</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Pre-test Occasion 1</th>
<th>Post-test Occasion 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$M$</td>
</tr>
<tr>
<td><strong>Embedded design</strong></td>
<td>68</td>
<td>80.39</td>
</tr>
<tr>
<td><strong>Applied experience</strong></td>
<td>20</td>
<td>76.65</td>
</tr>
</tbody>
</table>

N.B. 17 participants from the embedded design condition and 4 from the applied experience condition did not complete both pre and post-tests.

The participant numbers included for analysis here included the entire sample. A two-way Repeated Measures ANOVA was conducted with intervention condition (embedded design versus applied experience) as a between subjects factor and time (pre-test versus post-test) as a within subjects factor. The Condition x Time interaction was examined using the multivariate criterion of Wilks’s Lambda ($\Lambda$). A one-way repeated measures ANOVA was used to follow up a statistically significant interaction. Holm’s sequential Bonferroni procedure was used to control for Type 1 error across comparisons for simple effects that followed from the one way ANOVA. Alpha was set at the .05 level.

The self-efficacy scores on the SEIPD for Condition x Time interaction were NOT statistically significant, $\Lambda = .99$, $F(1,86) = 1.29$, $p = .26$, multivariate $\eta^2 = .015$. The
multivariate tests indicated a significant main effect for self-efficacy for time (pre versus post-test), \( \Lambda = .56, F(1,86) = 69.06, p = .00 \), multivariate \( \eta^2 = .45 \).

Figure 4.6 illustrates PST performance at each occasion by intervention condition. Data in both the table and the figure highlight the similarities between groups for occasion 2 and statistically did not favour either condition in terms of improvements in self-efficacy. Overall, the results indicated that participation in an inclusive education course did covary with stronger self-efficacy among pre-service teachers in both conditions. The scores under both conditions increased from pre- to post-test, indicating that self-efficacy levels improved irrespective of treatment condition. They did improve more for the applied experience condition than the embedded design condition.

![Figure 4.6 SEIPD mean scores by condition](image)

**Figure 4.6 SEIPD mean scores by condition**

**4.6 Study two results**

It is not possible to determine statistically significant differences for study two due to the small participant sample size. The research hypotheses for study two were as follows:

Study two: \( H_2 = \) Descriptive statistics for observation of classroom implementation will show higher levels of:
a) The frequency and sophistication of pattern language for PSTs who have experienced *embedded* course design when implementing ET or CL compared to those who experienced the *applied experience* course design;

b) Higher levels of self-efficacy between the PSTs who experience the application of *embedded design* principle to the theoretical design of the inclusive education course as opposed to those who participate in an *applied experience* condition

c) Design integrity for PSTs who have experienced *embedded* course design when implementing ET or CL compared to those who experienced the *applied* course design

d) Implementation integrity for PSTs who have experienced *embedded* course design when implementing ET or CL compared to those who experienced the *applied experience* course design; and

e) Self-awareness of PSTs about lesson design and implementation for those who have experienced *embedded* course design compared to those who experienced the *applied experience* course design.

\(H_{02} = \) Descriptive statistics for the observation of classroom implementation will show there will be no difference between:

a) The frequency and sophistication of pattern language for PSTs who have experienced *embedded* course design when implementing ET or CL compared to those who experienced the *applied experience* course design;

b) Higher levels of self-efficacy between the PSTs who experience the application of *embedded design* principle to the theoretical design of the inclusive education course as opposed to those who participate in an *applied experience* condition

c) Design integrity for PSTs who have experienced *embedded* course design when implementing ET or CL compared to those who experienced the *applied* course design

d) Implementation integrity for PSTs who have experienced *embedded* course design when implementing ET or CL compared to those who experienced the *applied experience* course design; and

e) Self-awareness of PSTs about lesson design and implementation for those who have experienced *embedded* course design compared to those who experienced the *applied experience* course design.
Mastery of course content (PCK) was not measured in study two. The following results present findings for the measures of: pattern language, self-efficacy, and implementation fidelity in the field. The data for all of the following dependent measures relates to 30 volunteers who participated in study 2.

4.6.1 Frequency of pattern language.

In study two, pattern language frequency was determined in the same manner as in study one. Rather than analyzing a reflection PSTs completed about their learning, in study two, lexicon specific to the characteristics of explicit teaching were counted based on a reflection PSTs wrote following the delivery of their lesson which occurred in university session 5 out of 8 of the PST program. The PSTs were required to write a reflection on “how they thought their lesson went”, with no other directions given. The results for pattern language frequency between the two conditions (embedded design and applied experience) are shown in Table 4.7. Comparisons can be made based on matched samples of participants across study one and two.

Table 4.7

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>PL Pre-test</th>
<th>M</th>
<th>SD</th>
<th>PL Post-test</th>
<th>M</th>
<th>SD</th>
<th>PL Post-practicum</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded design</td>
<td>20</td>
<td>1.48</td>
<td>1.29</td>
<td></td>
<td>28.14</td>
<td>4.63</td>
<td></td>
<td>6.30</td>
<td>1.92</td>
<td></td>
</tr>
<tr>
<td>Applied experience</td>
<td>10</td>
<td>1.90</td>
<td>1.20</td>
<td></td>
<td>23.4</td>
<td>6.88</td>
<td></td>
<td>2.56</td>
<td>3.43</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.7 depicts the frequency of pattern language used by PSTs when reflecting on their planning and delivery of an explicit teaching lesson. Pre-service teachers in the embedded design condition used pattern language 2.46 times more frequently than did students in the applied experience condition at the post-practicum occasion. The average mean of 5.14 across groups at post-practicum occasion demonstrates the ability of the PSTs to maintain the use of the pattern language of explicit teaching across time in both conditions. The post-prac results were at a much lower rate than they had been at the post-test occasion, but were not as low as in the original pre-test
for both conditions. It is acknowledged that the circumstances for post-prac data collection were different to those in the pre and post-test circumstances, however, these results support hypothesis H₃d, where descriptive statistics showed higher levels of pattern language frequency for participants in the embedded design condition than those in the applied experience condition.

4.6.2 Sophistication of pattern language.
Pattern language sophistication was examined using SOLO taxonomy. Pre-service teacher responses were evaluated in the same way as in study one where the levels of SOLO (Biggs & Collis, 1982) were applied to the reflection written by the PSTs following the delivery of their lessons.

Table 4.8 depicts the level of sophistication of pattern language used by PSTs when reflecting on their planning and delivery of an explicit teaching lesson. Pre-service teachers in the embedded design condition used pattern language in a more sophisticated manner (M = 3.9) than those in the applied experience condition (M = 2.22).

Table 4.9
Pattern Language Sophistication Mean and Standard Deviation Scores by Condition for Matched Sample of Participants

<table>
<thead>
<tr>
<th>Condition</th>
<th>SOLO ET quiz study 1</th>
<th>SOLO Post-practicum</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Embedded design</td>
<td>20</td>
<td>2.60</td>
</tr>
<tr>
<td>Applied experience</td>
<td>10</td>
<td>2.30</td>
</tr>
</tbody>
</table>

Overall, 50% of respondents were achieving close to level 4 SOLO responses incorporating pattern language lexicon into their reflections.

SOLO levels for ET quiz study 1 were taken when participants were undertaking a university course. The results of this quiz were relevant to study two comparisons as it was the quiz that focused specifically on explicit teaching and all of the observed lessons used explicit teaching as the pedagogy of choice. A procedure of matching
occurred so the same participants are being compared across time for each condition. The average SOLO level taken across quiz scores for the embedded design condition in *ET quiz study 1* was 2.56. In the post-practicum data point, the score was 3.9. The average SOLO level taken for the applied experience condition in *ET quiz study 1* was 2.26; at the post-practicum data point it was 2.22. The practical significance of these results indicate that those in the embedded design condition improved in the sophistication of their responses while those in the applied experience condition declined.

In summary, then hypothesis statement H$_2$d is supported for study two where descriptive statistics for sophistication of pattern language terms showed higher levels of sophistication for participants in the embedded design condition than those in the applied experience condition.

4.7 Self-efficacy

Table 4.9 indicates the results for matched samples of participants from the post occasion data collection point in session 3 of the university program to the post-practicum data collection point in session 5 of the university program for PSTs.

<table>
<thead>
<tr>
<th>Table 4.10</th>
<th>Self-efficacy (SEIPD) Mean and Standard Deviation Scores by Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>Post-test Session 3</td>
</tr>
<tr>
<td>Embedded design</td>
<td>N= 20</td>
</tr>
<tr>
<td>Applied experience</td>
<td>N= 10</td>
</tr>
</tbody>
</table>

The results could only be reported in terms of descriptive statistics for means and standard deviations due to the small sample sizes involved. The results indicated that levels of self-efficacy remained high for all participants for the ensuing 12 months following the completion of the inclusive education course and following weeks of in-school practicum experiences. Even though the results for those in the applied experience cohort improved from during data collection times, they were not at a practically significant level. These results support the null hypothesis: H$_{02}$e as the descriptive statistics did not show higher levels of self-efficacy for PSTs who
experienced *embedded* course design than those who experienced the *applied experience* condition.

4.8 Application of the Embedded Design Principle in Practice Teaching Settings

The independent variables for study two remained the same as in study one: that of instructional condition (*embedded design* or *applied experience*). The research questions specifically from study two sought to examine the effect of each intervention condition (*embedded design* and *applied experience*) on PST design fidelity, implementation fidelity, and then self-reflection of research-based pedagogies whilst on practicum during session 6 of their university program.

These hypotheses sought to establish if: a) pre-service teachers could generalise characteristics of explicit teaching in lesson design fidelity, delivery fidelity, and self-reflection when in an applied setting; and b) there was a differential effect of intervention condition on PST performance on measures of integrity of explicit teaching. There was no parametric analysis carried out for study two as the numbers of participants were significantly lower (\(N = 30\)). Instead, descriptive statistics and observations were relied upon for data collection and analysis.

4.8.1 Dependent variables.

The dependent variables for study two included: *use and sophistication of pattern language, levels of PST self-efficacy, design integrity, implementation integrity and self-reflection* of explicit teaching characteristics. The analysis of pattern language frequency for study two was reported in Table 4.7. Pattern language sophistication for study two was reported in Table 4.8 and the analysis of self-efficacy in study two at this observation time was reported in Table 4.9. The following discussion reports results of fidelity data while PSTs were on practicum.

Design, implementation, and self-reflection integrity referred to the ability of PSTs to design, deliver, and reflect on instruction using research-based characteristics of explicit teaching they had previously been studying in classes at university. Twelve months on from the completion of the inclusive education course at university, volunteers were sought for lesson observations during their third year (session 6) practicum placements.
Data were collected using a fidelity observation protocol devised by the researcher to reflect the characteristics of explicit teaching lessons as specified during university coursework. The protocol was used as a means of data collection for: (i) document analysis for design integrity; (ii) researcher observation of implementation fidelity and (iii) PST self-assessment of their design and implementation; with the same protocol being used in each of these three instances. Document analysis was used to determine design fidelity and involved the researcher scoring the written lesson plan provided by the PST and noting characteristics on the scoring protocol. Researcher implementation observations involved the researcher observing a lesson designated by the PST and noting characteristics as observed on the scoring protocol. The self-assessment used the same protocol and involved the PST scoring the same lesson that the researcher had observed after the lesson was completed.

During the researcher observations, inter-observer reliability was assessed between the researcher and a volunteer observer. Twenty percent of lessons were observed for reliability purposes and a percentage of inter-observer agreement was calculated at 95% across observations. Reliability was augmented by comparisons made between the data sources as well: The actual observation, the self-reporting from the PST of the lesson, and the written documentation (the design of the lesson plan) of the observed lesson. Design, implementation and then self-evaluation results will now be elaborated.

Table 4.10 provides a summary of the raw data for these dependent variables. Detailed analysis of each sub element of explicit teaching strategies follows in Table 4.10 for: (i) Implementation fidelity; (ii) student self-assessment fidelity; and (iii) document analysis for design fidelity. Actual numbers in each condition are provided as well as percentages to enable comparison. Each of the variables will be elaborated in the sections following the table.
Table 4.11
Implementation Fidelity, Self-reflection by PSTs and Design Fidelity

<table>
<thead>
<tr>
<th>Lesson characteristics</th>
<th>Implementation fidelity</th>
<th>Self-reflection fidelity</th>
<th>Design fidelity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ED (N)</td>
<td>AE (N)</td>
<td>ED (N)</td>
</tr>
<tr>
<td></td>
<td>% correct</td>
<td>% correct</td>
<td>% correct</td>
</tr>
<tr>
<td>Learner outcome for ET lesson</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition:</td>
<td>(15)75</td>
<td>(2)20</td>
<td>(17)85</td>
</tr>
<tr>
<td>Action:</td>
<td>(15)75</td>
<td>(2)20</td>
<td>(15)75</td>
</tr>
<tr>
<td>Criterion:</td>
<td>(6)30</td>
<td>(0)0</td>
<td>(13)65</td>
</tr>
<tr>
<td>Key words</td>
<td>(20)100</td>
<td>(10)100</td>
<td>(19)95</td>
</tr>
<tr>
<td>The attention grabber</td>
<td>(20)100</td>
<td>(8)80</td>
<td>(17)85</td>
</tr>
<tr>
<td>Link to prior learning</td>
<td>(19)95</td>
<td>(9)90</td>
<td>(19)95</td>
</tr>
<tr>
<td>Model</td>
<td>(14)70</td>
<td>(6)60</td>
<td>(13)65</td>
</tr>
<tr>
<td>Guided practice</td>
<td></td>
<td></td>
<td>(6)60</td>
</tr>
<tr>
<td>GP is broken down into logical steps</td>
<td>(15)75</td>
<td>(2)20</td>
<td>(10)50</td>
</tr>
<tr>
<td>GP relates to the model</td>
<td>(11)55</td>
<td>(3)30</td>
<td>(13)65</td>
</tr>
<tr>
<td>Teacher rotates</td>
<td>(16)80</td>
<td>(4)40</td>
<td>(14)707</td>
</tr>
<tr>
<td>Independent practice</td>
<td>(17)85</td>
<td>(3)30</td>
<td>(9)45</td>
</tr>
<tr>
<td>ET differentiation:</td>
<td></td>
<td></td>
<td>(4)40</td>
</tr>
<tr>
<td>Adapt a learner outcome</td>
<td>N=13</td>
<td>N=2</td>
<td>N=15</td>
</tr>
<tr>
<td>Differentiate the condition, or</td>
<td>(2)15</td>
<td>(0)0</td>
<td>(7)46.7</td>
</tr>
<tr>
<td>the product or action, or</td>
<td>(7)51</td>
<td>(1)50</td>
<td>(10)66.7</td>
</tr>
<tr>
<td>the criteria</td>
<td>(7)51</td>
<td>(1)50</td>
<td>(12)80</td>
</tr>
</tbody>
</table>
4.8.2 Lesson Design Fidelity.
When scoring design fidelity, the characteristic of criteria scored lowest of the three elements in the learner outcome. Percentages correct for the three elements of the learner outcome in the embedded design condition remained comparable to researcher observation scores and self-assessment scores. In the applied experience condition, however, there was zero evidence of the learner outcome characteristics. Participants in the applied experience condition did not attempt any elements of the learner outcome. Instead of noting specific condition, actions and criteria in their documentation, they included outcomes taken directly from the syllabus.

Scores for both conditions remained similar for the characteristics of: key words, attention grabber, link to prior learning and model. The characteristics of guided practice and independent practice indicated scores for applied experience that were noticeably lower in all counts. For example: GP broken down into logical steps (ED=75%, n=8; AE=10%, n=1); GP relating to the teacher modelling (ED=58.3%, n=8; AE=10%, n=1); and teacher rotating to check understanding (ED=58.3%, n=8; AE=20%, n=2) all higher than the applied experience condition.

Within the differentiation characteristics of explicit teaching, participants in the embedded design condition wrote the characteristics of differentiated content, product and criteria in their lesson plan documentation. Pre-service teachers in the applied experience did not attempt to include differentiation possibilities in their written lesson planning, even though one of these participants was able to demonstrate differentiation of product and criteria in their lesson.

The fidelity observation protocol surveyed critical elements in the design and delivery of explicit teaching lessons. For example, on the explicit teaching protocol (11 core items plus 3 more for differentiation) the observer was required to determine whether characteristics were present, partially present, or not present. The observer recorded the presence or absence of each characteristic along a continuum using a Likert scale ranging from 0 to 2. Essential characteristics of practice were allocated a score of: complete = 2; partially complete = 1 and incomplete = 0. A total score was also computed, where the maximum obtainable was 24. Scores were discussed in terms of percent correct for ease of comparison (see Appendix 6 for scoring protocol). The
results presented here are items where the participants scored the maximum mark possible for that characteristic (i.e., only those who scored 2/2 were reported as correct). Scores of 0/2 (not present) and 1/2 (partially present) were omitted from this analysis and discussion.

Table 4.11 provides descriptive statistics using scores for document analysis of design fidelity. The mean scores for the embedded design condition were 9.46 higher than for the applied experience condition. The PST educators in the embedded design condition were more often able to demonstrate essential characteristics of explicit teaching in written lesson design documentation than those of the applied experience condition.

Table 4.12
Design Fidelity by Instructional Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded design</td>
<td>14</td>
<td>15.36</td>
<td>4.52</td>
</tr>
<tr>
<td>Applied experience</td>
<td>10</td>
<td>5.90</td>
<td>3.23</td>
</tr>
</tbody>
</table>

N.B. 6 participants from the embedded design condition did not complete a lesson plan for analysis.

The results shown in Table 4.11 support hypothesis H2a which stated that participants in the embedded design condition would have a higher planning and design integrity when planning classroom lessons than those in the applied experience condition. That is, they were able to incorporate more essential explicit teaching characteristics in their lesson plan designs. Results show that PSTs in the embedded design condition were aware of the characteristics they needed to be using when designing the lesson.

4.8.3 Implementation fidelity.

Considering researcher implementation fidelity scores, the first three scale items in Table 4.10 (condition, action and criterion) constituted the core items associated with writing a learner outcome for an explicit teaching lesson. In all instances, participants in the embedded design condition wrote learner outcomes with a significantly higher level of fidelity than those in the applied experience condition. For example, in the ‘condition’ section of learner outcomes, the embedded design condition was 50 percentage points higher when preparing learner outcomes (ED=15, AE=2). For the ‘action’ component of learner outcomes, embedded design was 50 percentage points higher (ED=15, AE=2), and for the criteria’ component, embedded design only scored
28% while the *applied experience* condition scored zero for this component of learner outcomes (ED=6, AE=0). Perhaps this is linked to the pedagogical content knowledge results achieved when participants were studying the university course 12 months earlier. At post test for study one, participants in the *embedded design* condition scored mastery level of pedagogical content knowledge scores significantly higher than participants in the *applied experience* condition.

Participants in both conditions were implementing at a level of 80 percentage points or higher for characteristics of noting *key words* (ED, n=21; AE, n=10) establishing an *attention grabber* (ED=20, AE=8) and making a *link to prior learning* (ED, n=19; AE, n=9). This high level of implementation across conditions reflects the notion that these characteristics were easier for the PSTs to generalise to an applied setting.

Participants in the *embedded design* and *applied experience* conditions scored 66.7% (N=14) and 60% (N=6) correct respectively for *teacher modelling* required teaching the focus learner outcome. This result indicates that both groups had not reached mastery in terms of linking their model directly back to the action in their learner outcome. *Pedagogical content knowledge* of a particular characteristic is one level of ability; the demonstration of that characteristic is more difficult.

*Guided practice* again provided a separation of the groups where *embedded design* was significantly higher for breaking down the model into *logical steps* (50 percentage points higher, ED=15, AE=2); *relating the steps directly to the teacher model* (22 percentage points higher, ED=11, AE=3); and *checking for understanding* (36 percentage points higher, ED=16, AE=4) on each of the steps. *Independent practice* also favoured the *embedded design* condition, scoring 51 percentage points higher than the *applied experience* condition (ED=17, AE=3).

Scores for *differentiation of the learner outcome* were based on a much lower number of participants, 13 of the 21 participants (62%) attempted the differentiation of *condition, product or content* from the *embedded design* condition and 2 of the 10 participants (20%) attempted these elements from the *applied experience* condition. Of those from the *embedded design* condition, 19% (ED=2) were correct in their differentiation of the teaching *conditions*, 57% (ED=7) were correct with differentiation of the *product* of action, and 57% (ED=7) were correct with the *criteria*
differentiation. Of those from the \textit{applied experience} condition, none were successful with differentiation of the teaching \textit{condition}, and 50\% (AE=1) were successful with differentiation of the \textit{product} and \textit{criteria} of the learner outcome; however these percentages should be interpreted with caution when only two participants were involved.

Overall scoring was lower for the \textit{embedded design} condition for differentiation than for the first 11 scale characteristics; however 62\% (ED=13) did attempt differentiation for the students concerned. The fact that 80\% (AE=8) did not even attempt to differentiate from the \textit{applied experience} condition is perhaps the result of very low scores for the previous 11 characteristics of explicit teaching. Pedagogically speaking, these would need to be in place and correct before differentiation should be attempted. Each condition scored lower on differentiation of the \textit{condition}, and chose to differentiate the \textit{product} or the \textit{criteria}. Perhaps this is an indication that differentiation of the latter two is an easier task to consider initially. This discussion again supports the hypothesis $H_2b$ where higher levels of implementation integrity were evident for participants in the \textit{embedded design} cohort.

Table 4.12 provides descriptive statistics for researcher observation of \textit{implementation fidelity} using raw observation scores.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|}
\hline
\textbf{Condition} & \textbf{$N$} & \textbf{$M$} & \textbf{$SD$} \\
\hline
\textit{Embedded design} & 20 & 16.57 & 3.92 \\
\textit{Applied experience} & 10 & 9.00 & 3.43 \\
\hline
\end{tabular}
\caption{Implementation Fidelity by Instructional Condition}
\end{table}

Table 4.13 results support the hypothesis $H_2b$ that students in the \textit{embedded design} condition actually implement explicit teaching lessons with a higher degree of implementation fidelity than those PSTs in the \textit{applied experience} condition. That is, they incorporated more essential characteristics of explicit teaching in their teaching practice.
4.8.4 Self-evaluation fidelity.

Considering the self-evaluation scoring, participants in the applied experience condition scored 100% for sections of key words and attention grabber. Participants in the embedded design condition scored themselves higher on all other essential characteristics of explicit teaching. The differentiation scores are difficult to give credence to when only two completed the self-assessment in the applied experience condition. Pre-service teachers in the embedded design condition scored themselves higher than the researcher observing the same lesson for 5/14 of the scale items. Pre-service teachers from the applied experience condition scored themselves higher than the researcher on 6/14 scale items. This indicates that participants in both conditions are not aware of the errors in implementation they are making. If PSTs do not have an accurate perception of their own behaviour then that could conceivably influence their self-efficacy as well. There may be potential links to facility with concepts, maintenance of skills, and course content/delivery issues that need to be investigated following this result.

For PSTs in the embedded design condition, one of the largest differences between researcher observation and the self-scoring was on the scale of criterion. For example, the researcher scored (ED=8) 28% correct for criterion where the PSTs scored (ED=13) 61.9% correct for the same characteristic. Clearly the PSTs thought they were addressing the need to have criteria explicitly embedded in their lesson plan, where the researcher did not observe this characteristic.

Pre-service teachers also scored themselves quite low on the independent step of the explicit teaching characteristics. Based on verbal comments made by the PSTs, this may have been due to many of the school students not finishing the work that had been set for them, so they never reached the independent stage of the lesson.

PSTs in the applied experience condition scored much closer to the researcher observations with the self-assessment (for example: the scores for ‘model, and GP related to the model’ were the same between PSTs and researcher. ‘GP broken down…, teacher rotating, and independent practice’ had only 10 points difference between PSTs and researcher).

Perhaps PSTs in the applied experience condition recognised their incorrect application of the characteristics. The hypothesis H2c is supported where descriptive statistics showed higher levels of self-awareness scoring from PSTs who experienced the embedded design condition compared to those who experienced the applied experience condition.
These results support the hypothesis H2a where descriptive statistics showed higher levels design integrity for PSTs who experienced the *embedded design* course compared to those who experienced the *applied experience* course design. The results also support hypothesis H2b where descriptive statistics showed higher levels implementation integrity for PSTs who experienced the *embedded design* course compared to those who experienced the *applied experience* course design which is now discussed.

Table 4.13 provides descriptive statistics for PST *self-evaluation* of their implementation fidelity using raw scores for study two. The mean scores for PSTs in the *embedded design* condition (17.05) were higher than for those in the *applied experience* condition (12.90) during researcher observations, with a difference in mean scores of 7.57 in favour of the *embedded design* condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded design</td>
<td>20</td>
<td>17.05</td>
<td>3.75</td>
</tr>
<tr>
<td>Applied experience</td>
<td>10</td>
<td>12.90</td>
<td>3.45</td>
</tr>
</tbody>
</table>

The mean score for the *embedded design* condition was 4.15 higher than for the *applied experience* condition. The PSTs in the *embedded design* condition perceived themselves to be incorporating essential characteristics of explicit teaching while implementing a lesson in a school setting at a higher level than those in the *applied experience* condition.

Table 4.13 results support the hypothesis H2c that students in the *embedded design* condition self-reflected on their explicit teaching lessons with a higher degree of fidelity than those PSTs in the *applied experience* condition. That is, they were aware of more essential characteristics of explicit teaching into their teaching practice.

### 4.9 Summary

This chapter has presented the results from study one and two. A summary of answers to the hypotheses asked at the beginning of this chapter are as follows. In study one,
there was a statistically significant difference between the PSTs who experienced application of the *embedded design* principle of an inclusive education course as opposed to those who participated in an *applied experience* condition in:

- \( H_1a \). The mastery of course content in the form of pedagogical content knowledge (PCK) and
- \( H_1b \). The frequency of pattern language.

The null hypotheses \( H_{01} b \) demonstrated no statistically significant difference in the levels of sophistication of pattern language and also \( H_{01} c \) no statistically significant difference in the levels of Self-efficacy between the PSTs who experienced application of the *embedded design* principle to the theoretical design of an inclusive education course as opposed to those who participated in an *applied experience* condition.

The research hypotheses for study two indicated that descriptive statistics for classroom implementation showed higher levels of the following for PSTs who have experienced *embedded* course design when implementing explicit teaching compared to those who experienced the *applied* experience course design for:

- \( H_2 a \). Design integrity of research based pedagogies
- \( H_2 b \). Implementation integrity
- \( H_2 c \). Self-awareness of PSTs about lesson design and implementation and
- \( H_2 d \). The frequency of pattern language

Descriptive statistics for the observation of classroom implementation showed there was no difference between \( H_{02} e \). Self-efficacy between the PSTs who have experienced *embedded* course design when implementing ET compared to those who experienced the *applied* experience course design. The following chapter (Chapter 5) will discuss these findings in relation to the literature and follow up with directions for future research.
Chapter 5: Discussion

All students depend on highly qualified teachers to carry out efficacious interventions (Hattie, 2009). These interventions are required by teachers to not only meet the needs of their students, but also to deal with change and solve problems that arise in complex inclusive learning environments (Forlin et al., 2015; King-Sears et al., 2004).

Pre-service teacher (PST) education programs are the vehicles for providing teachers with preparation for inclusive classrooms. Discussions in the literature surrounding teacher preparation for inclusion are extensive (Jung, 2007; Killoran et al., 2014; Romi & Leyser, 2006). While there is widespread support for university-based teacher preparation, there continues to be national and international concern about whether the preparation PSTs receive for inclusion is adequate, and whether it does what schools require to ensure successful inclusive education (Carroll et al., 2003; Edelen-Smith et al., 1993; Forlin et al., 2015; Husebo, 2012; Snyder, 2012).

A major study into Australian teacher preparation (Ramsey, 2000) found that evidence-based research was promoted as critical to the field and as a key way to identify and validate effective models of teacher education. However, little of this type of research is actually carried out (Burkhardt & Schoenfeld, 2003; Grima-Farrell, 2012; Hempenstall, 2006; Ure et al., 2009; Zundans-Fraser, 2014).

In a major study in the United States of America, Cochran-Smith and Zeichner (2005) developed a substantial evidence base of and recommendations for practices to be used in PST education. They noted that major requirements of PST programs should be to establish a shared language in teacher education curriculum to support collaboration, and to ensure competence in PSTs before graduation. These requirements noted by Cochran-Smith and Zeichner (2005) were echoed by Florian (2011) six years later. The knowledge required by inclusive teachers includes the skills of using high fidelity research-based pedagogies and collaboration to work productively with colleagues using sophisticated language.

By 2014 there had been several reviews of empirical studies about learning to teach. Yet, Sleeter (2014) still found that only 14% did something other than report or comment on existing research. Ludlow et al. (2010) proposed that the lack of empirical
research has amplified the debate about what kinds of evidence are appropriate and how different forms of evidence should be used for PST education.

Few studies have described the specific practices used in higher education to prepare PSTs to work with diverse populations of students (Jimenez-Silva & Olson, 2012) and, more importantly, their effect on PST professional learning. In general, there have been few empirical studies of PSTs learning to teach students with diverse needs and even fewer where PSTs learned the skills required of inclusive educators and could demonstrate those skills in applied settings (Florian, 2011).

The theoretical design rationale and content composition of pre-service courses and their impact on future inclusive practice is particularly important in this regard and requires further scrutiny (Billingsley et al., 2004; Darling-Hammond, 2005; Darling-Hammond et al., 2005; Fore et al., 2002). The review found no empirical studies specifically focused on the way a theoretically designed teacher preparation course could influence PST learning about inclusion, nor did any studies examine the extent to which that learning translated to classroom practice.

Given the importance of teacher preparation for inclusive education and the lack of research showing the impact of PST education on knowledge and skills, the intention of this research was to determine whether course design would have an impact on whether PSTs were able to:

1. learn pedagogical content knowledge (PCK) about research-based pedagogies that are known to have a positive impact on student outcomes;
2. utilise a professional lexicon (pattern language, PL) that is a necessary requisite for sophisticated collaboration about teaching;
3. foster a positive level self-efficacy (SE) that PSTs have about their ability to successfully teach students who have varying needs; and to
4. improve performance in these skills and translate this into practice in the classroom.

The review of the literature provided the context for and the basis for identification of these specific factors under investigation.
The following figure describes the sequence for establishing the differential impacts of the *embedded design* principle on skills, knowledges and attitudes of PSTs.

*Figure 5.1.* The application of course design to factors relevant to successful inclusive practice.

In this study, the *embedded design* condition noted in Figure 5.1 was derived from the application of a central theoretical principle from self-organizing systems and then applied to inclusive education course development. This condition was compared to a condition that incorporated experience in the field which is often linked to a greater praxis between theory and implementation of skills and knowledges (Bain, Lancaster, Zundans & Parkes, 2009; Hanline, 2010).

The following discussion is presented in three sections to cover each of the inclusive education factors: pedagogical content knowledge, use of sophisticated pattern language, levels of self-efficacy, design integrity of pedagogies, implementation integrity of pedagogies and self-reflection about those pedagogies. First, the research questions and hypotheses are revisited. Second, a brief summary of the research findings is presented. Third, the findings are discussed in terms of implications for PSTs.
5.1 Pedagogical Content Knowledge (PCK)

5.1.1 The research question.
Does application of the embedded design principle to the design of a PST inclusive education course result in improvements (overall and differential improvements) in the mastery of pedagogical content knowledge (PCK)?

5.1.2 Summary of research findings.
In study one, hypothesis $H_{1a}$ is supported. There was a statistically significant difference between the PSTs who experienced application of the embedded design principle of an inclusive education course as opposed to those who participated in an applied experience condition in mastery of course content in the form of pedagogical content knowledge (PCK).

The results showed that overall there were statistically significant improvements in PCK for both conditions between the pre- and post-tests. PSTs learned more at a statistically significantly level under the embedded design condition at post-test condition ($ED = 26.22$ and $AE = 18.82$) and, importantly, their levels of learning were at a threshold level designated to indicate mastery (80%) of the learning material. In the applied experience condition, PST students reached a mastery level of 80% in just one of the three quizzes.

On the three occasions where quizzes were administered to the groups, the embedded design condition scored higher than the applied experience group on each occasion. While there were not statistically significant differences between the conditions at quiz 1 and quiz 2, there was a statistically significant difference at quiz 3 with a mean for the embedded design cohort of more than twice that of the applied experience condition (quiz 3 $ED = 9.48$ and $AE = 4.14$). This finding supports the idea that the embedded design approach attenuates the achievement distribution by increasing overall PST mastery of PCK. In summary, the embedded design approach proved to be more robust in terms of PST knowledge than the applied experience approach.
5.1.3 Discussion of the research findings.

There have been repeated calls in the literature to improve the PCK of teachers (Cochran-Smith & Ziechner, 2005; Cook & Tankersley, 2012; Florian, 2011; Kretlow & Helf, 2013; Sleeter, 2014). Recommendation 6 from the Australian Institute for Teaching and School Leadership body (AITSL, 2015) states: Initial accreditation of programs require higher education providers to demonstrate that their programs have evidence-based pedagogical approaches, to ensure pre-service teachers are classroom ready, and that higher education providers provide a set of measures that assess the effectiveness of their programs in achieving successful graduate outcomes (2015). Also, recommendation 14: Higher education providers deliver evidence-based content focused on the depth of subject knowledge and range of pedagogical approaches that enable pre-service teachers to make a positive impact on the learning of all students (AITSL, 2015).

Pre-service teachers need to know about the key features of their professional practice if they are to act on their intentions and dispositions and develop teaching skills pertaining to that knowledge. The lack of empirical research, in the area has caused a number of authors to advocate for a more empirical evaluation of facilitators known to enhance inclusive practice (such as use of consistent language and research-based pedagogies). This thesis progresses the solution generation urged by Arthur-Kelly et al. (2013) by providing an empirical means of differentially evaluating pedagogical content knowledge of PSTs based on course design.

Of particular importance in this thesis was the finding of mastery level quiz performance in the embedded design condition. The finding of a mastery achievement distribution and low variance in PST scores in this study stands in contrast to the more common normal achievement distributions produced in the applied experience condition and by other courses in the PST program (as reported anecdotally by PSTs.) This demonstrates that PSTs attained a high level of understanding of the key concepts and ideas associated with the quiz content, creating a strong foundation for subsequent pattern language and skill development. It is important to note that mastery experiences consistent with those described here have also been identified as a source of self-efficacy related to working with children who have varying needs (Avery & Meyer, 2012; Holzberger et al., 2013). Yilmaz (2011) determined that the more proficient teachers perceived themselves to
be with PCK (for example: about explicit teaching and cooperative learning), the more efficacious they felt in the implementation of those instructional pedagogies.

These results show that embedded design was successful in building the base of knowledge for implementing the inclusive pedagogy addressed in the course with integrity (Klinger et al. 2004; Kretlow & Helf, 2013). These findings are consistent with prior research (Bain, Lancaster, Zundans & Parkes, 2009) who sought to establish whether the application of the embedded design principle covaried with the development of PCK in earlier studies using a within subjects design. In one such study, Bain, Lancaster, Zundans and Parkes (2009) found that structured group work provided results statistically higher than other forms of quiz preparation such as peer tutoring or individual learning.

After establishing support for the model by demonstrating the positive impact on student learning of PCK, the second step was to establish whether improved learning translated into the development and sustainability of professional pattern language required for the application of learning in practice. Pattern language will now be discussed first in terms of frequency, and following that, the sophistication with which that language was used.

5.2 Pattern Language

In this study, pattern language is the term used to describe the professional language that describes research-based characteristics of pedagogies that PSTs were learning about. PSTs’ learning was expressed in terms of the frequency of pattern language (lexicon) use as well as the sophistication of that use. The use of a common professional pattern language in collaboration with colleagues is important to inclusive educators as a means to solve problems about students, their needs and learning at a sophisticated level. A shared pattern language aids the ability to provide feedback about teaching approaches, student need and decision making (Friend & Cook, 2013; Hennessey & Dionigi, 2013).

5.2.1 The research questions.

Does application of the embedded design principle to the design to a PST inclusive education course result in improvements or differential effect on PST performance on measures of pattern language frequency and are any effects maintained across time?
Does application of the *embedded design* principle to the design to a PST inclusive education course result in a differential effect of intervention condition on *sophistication of pattern language*? Was the pattern language lexicon differentially maintained by the PSTs as determined by intervention conditions across time?

### 5.2.2 Summary of research findings.

In study one, hypothesis $H_{1b}$ was supported as there was a statistically significant difference between the PSTs who experienced application of the *embedded design* principle of an inclusive education course as opposed to those who participated in an *applied experience* condition for the frequency of pattern language. The null hypotheses ($H_{01b}$) demonstrated no statistically significant difference in the levels of sophistication of pattern language between the two cohorts.

In study two, the hypothesis $H_{2d}$ was supported where descriptive statistics for classroom observation showed higher levels of both the frequency and sophistication of pattern language for PSTs who experienced *embedded* course design when implementing ET compared to those who experienced the *applied* experience course design.

The frequency of pattern language use associated with PCK was determined by analysing pre- and post-intervention assessments and quiz reflections at weeks 4, 8 and 12 of the course. It was again assessed 12 months later when the PSTs were in schools on their practicum (six occasions in total).

The participants in both conditions improved in the frequency of pattern language across time from pre- to post-test occasions; and there was a statistically significant difference *between* the intervention conditions in favour of the *embedded design* condition at the time of the post-test. Improved pattern language frequency was also demonstrated by both conditions across quiz occasions (weeks 4, 8 and 12). The statistical analysis indicated statistically significant differences at each quiz occasion in favour of the *embedded design* condition.

When the frequency scores were analysed following a further 12 months of PST program of study and completion of a practicum experience (study two). PSTs in the
*embedded design* condition used pattern language 2.46 times more frequently than PSTs in the *applied experience* condition at this final stage of the study. The overall use of pattern language words in the post-lesson reflection across groups demonstrates the ability of the PSTs to sustain the pattern language across time in both conditions.

Using the SOLO taxonomy, improvements in the sophistication of pattern language use were assessed at quiz occasion for each condition (*embedded design* and *applied experience*) occurring during weeks 4, 8 and 12 of the course. For both conditions, results indicate a statistically significant effect over time, with follow-up comparisons indicating statistically significant differences between quiz 1 and quiz 2. There was also a statistically significant difference between quiz 2 and 3, but in a negative direction for both conditions. That is, even though statistically significant differences were maintained between the conditions for quiz 2 and 3, both declined in overall results due to the cumulative difficulty of the questions involved. Slightly higher means were achieved at each quiz occasion for the *embedded design condition* even though the differences were not statistically significant.

When analysed 12 months later, whilst the PSTs were on practicum, those in the *embedded design* condition used pattern language in a more sophisticated manner than those in the *applied experience* condition. Students in the *embedded design condition* increased from one year to the next with more than half of the responses at this latter time using language at the level four (Mean of 3.9) of the SOLO taxonomy (Biggs & Collis, 1982). This level is known as the *relational level* and involves the cohesive linking of various pieces of information. PSTs in the *applied experience* condition declined in the level of sophistication to level two (Mean 2.22) of the SOLO taxonomy for the language they were using whilst on practicum. This level is known as the *unistructural level* and involves providing only one piece of relevant information, or one fact.

### 5.2.3 Discussion of the research findings about pattern language

A number of studies have described the importance of being able to communicate and collaborate about student learning using professional language (Hennessey & Dionigi, 2013; Jenkins et al., 2003; McLaughlin & Talbert, 2001). Pattern language is required by teachers as a basis, and then as the means, to collaborate at a sophisticated level with others about relevant pedagogies.
The results of this study build on the earlier (a within subject design) study by Bain, Lancaster, and Zundans (2009) that found mastery of pattern language frequency and sophistication covaried with participation in a pre-service course, and increased over time. There were no studies that directly measured frequency and sophistication of pattern language in a comparative manner. For example, the results found here are direct measures of pattern language frequency and sophistication.

The most obvious finding of the comparison between cohorts in this study was that the PSTs in both conditions did increase their capacity in terms of the frequency of pattern language related to inclusive education over the course of the study. The increase in pattern language frequency was statistically significantly greater for the embedded design condition at the post-test occasion. Both conditions enabled PSTs to increase their use of appropriate pattern language associated with research-based teaching pedagogies they might use in the classroom. The structure of the language of the embedded design condition was statistically significant and more effective at every point of frequency data collection for the non-comparative group design.

The findings for research question two are consistent with the acquisition of mastery level PCK by participants in the embedded design group, and may explain their more frequent and sophisticated use of pattern language compared to those in the applied experience condition. By this, it was indicated that PSTs mastered PCK at a higher level; and the higher level in the embedded design condition may have influenced the frequency of their professional language use.

The scores for frequency were lower during the practicum point in time than at the post-test occasion that occurred 12 months earlier. Embedded design went from an average mean frequency score of 7.93 to 6.3 and the applied experience design went from a mean of 4.58 to a mean of 2.56. In the interim, PSTs in both conditions had no ongoing engagement with the content of the course and as such a reduction could be expected as a ‘washout effect’ given their lack of continued use of the information and concepts. This ‘washout effect’ was less for the embedded design condition than for the applied experience. The particular “inclusive education pedagogical pattern language” addressed in the course was not supported nor developed in other courses from the PST program being completed in the interim timeframe. Both conditions
were able to maintain some level of pattern language usage when describing their practice.

Pattern language is seen as a prerequisite for inclusion if teachers are to be able to discuss pedagogies at the accurate and sophisticated level called for in the literature (Bain, Lancaster & Zundans, 2009; Hennessey & Dionigi, 2013; Jenkins et al., 2003). Hennessey and Dionigi (2013) used self-report to determine sophisticated understandings of pedagogy; Jenkins et al. (2003) used interviews to gather information. Both studies involved practicing teachers in their samples and showed that participants possessed a low-level understanding of important pedagogical characteristics. Even where a particular pedagogy was seen to be ‘embraced’ by the teachers, essential characteristics of the pedagogy were missing in their discussion. If important pedagogical characteristics are omitted, the strategy becomes less efficacious for school student achievement.

In this study, improvements in sophistication of pattern language were evident for both intervention conditions from quiz 1 to quiz 2. The sophistication scores both fell between quiz 2 and quiz 3 in equal proportions for both conditions. This result stands in contrast to the findings of mastery for the PCK of PSTs where one could suggest that with additional content being covered there would be more to learn and discuss. These results may reflect the difficulty of the content involved in quiz 3 and its scope. The content covered in quiz 1 focused on explicit teaching and in quiz 2 focused on cooperative learning. Quiz 3 was a continuation of these topics and included methods of differentiating both explicit teaching and cooperative learning.

The difficulty of the curriculum content being covered needs to be taken into consideration when PSTs are expected to learn the associated lexicon. Initial content covered pedagogical characteristics of both explicit teaching and then of cooperative learning. The third cycle of content delivery involved an extension of these initial characteristics to include differentiation of process, product, or content to a least three different levels. It is known from the PCK data gathered that the pattern language terms and associated concepts were learnt to a mastery level of 80% correct. Perhaps this expectation of mastery needs further time for maintenance and further investigation when delivering complicated curriculum and differentiation to PSTs. The programs of PST education also need to be analysed to determine levels of coherence.
where skills are developed and then revisited and practiced throughout the program in order to maintain mastery levels of knowledge and language. This had not been the case in the program studied.

The findings suggest a link between the frequency with which a lexicon is used and the sophistication of use. This link requires further investigation in terms of the mastery actually required to allow a flow on to sophistication. As with any use of explicit teaching, perhaps the sophistication of lexicon needs to be explicitly taught (Rosenshine, 2012) and not assumed to simply generalise just because frequency has been mastered at one point in time. At this point, we have evidence of the benefits of the embedded design for PST PCK learning, pattern language frequency, and now for the sophistication as well.

Slightly higher means were achieved at each quiz occasion for the embedded design condition when compared to the applied experience condition, even though the differences were not statistically significant. The quizzes held during weeks 4, 8 and 12 suggest that the embedded design group was more effective at a very early stage in the sequence of the course (i.e., after only four weeks of classes) and maintained effectiveness throughout the course. There was no empirical body of work found that directly measured frequency and sophistication of pattern language used by PSTs when discussing pedagogies.

The results here are an extension of those found by Bain, Lancaster and Zundans, (2009), where the largest positive effect occurred with the groups relying on collaboration to discuss the pattern language specific to their lesson designs. By collaborating with peers as part of the embedded design condition and using the required pattern language in an authentic situation such as this, it may have enhanced the understanding and thus sophistication of the interactions used by PSTs. The results also lend support to the findings from Waterman et al. (2010) who found the application of structured collaboration resulted in enhanced PST learning when compared with groups where no structured collaboration was in place.

Wenger, McDermot, and Snyder’s notion that “communities … involve those who wish to deepen their knowledge and expertise about a shared concern, process or problem through ongoing interaction” (2002, p. 95) helps to explain the differential
results. The ongoing interaction in this instance was enabled through the use of *embedded design* throughout weekly meetings, where the structured problem-solving process was embedded in and practiced by tutorial groups. The problem-solving characteristics of mutual engagement, joint enterprise and a shared professional repertoire were encouraged in the weekly preparation by PSTs to share their understandings about their newly learned lesson pedagogies by way of pattern language. The structured weekly collaborative problem-solving processes appear to have been more beneficial to PSTs in terms of pattern language sophistication than the field-based placements. In the *applied experience* condition, five of the weekly two-hour meetings/tutorials were replaced by the visits to work with school students where the structured form of collaborative problem solving could not be assured. While pattern language may have been a prerequisite to collaborative problem solving, the process of collaboration was not an independent variable in this study. It is discussed here as a possible explanation for the findings and as a useful consideration for future research.

The sophistication of inclusive education lexicon used by PSTs in the *embedded design* group (Mean = 2.6) occurred predominantly within the second unistructural category (SOLO 2), indicating mostly one piece of information was evident in a response, the multi-structural category (SOLO 3) and some within the relational category (SOLO 4), indicating that these PSTs were able to present multiple professional ideas, and in the relational responses link those to a main idea or concept. PSTs in the *applied experience* condition scored an average mean (2.36) in the second unistructural category (SOLO 2) indicating mostly one piece of information was evident in a response or that responses usually contained only one fact. Twelve months later whilst on practicum, PSTs in the *embedded design* group improved on their earlier results and scored a SOLO mean of 3.9 and those in the *applied experience* group decreased their results and scored a SOLO mean of 2.22.

Few responses fell into the extended abstract category (SOLO 5), which requires evidence of generalisation beyond the immediate context. This level of response is consistent with the kind of schema development where pattern language terms are framed by and contribute to a broader and deeper conceptual understanding of practice (Bain, Lancaster, & Zundans, 2009). Using this notion of conceptual understanding, few PSTs appeared to respond in a manner that was indicative of the existence of a
broader schema at this early stage of their undergraduate teacher preparation. The highest SOLO level (where additional information is provided by the PSTs that goes beyond what is expected in the given context) was not expected from these participants as this course was a basic introductory course within their undergraduate program rather than at the postgraduate level where a more expansive and sophisticated response could be expected.

In terms of the course design framework, this may indicate the necessity for PSTs to be engaged in courses utilising the same design across the PST program rather than just a single instance of exposure in one course. Such a schema could be expected to emerge from a cumulative and interrelated professional exposure across an entire teacher preparation program. As such, extensive schema development was not expected as an outcome of the experience in just one course. What is clear, however, is that the pre-requisite pattern language required to build that schema was developed by PSTs with the introduction and accumulation of the teaching cycles in both cohorts and maintained over time. Even teachers at the pre-service level of their training were able to share expertise and bring stakeholders together in a problem-solving endeavour (Buysse et al., 2003). The results show that PSTs can, as a result of their course experience, build a capacity to develop a professional language and express it in their professional work and in reflections about that work.

Using simulated settings, Leaman and Flanagan (2013) used role plays to analyse PST commentaries on lessons in terms of the meta-language used. This was a direct measure of their language used to describe the pedagogies. In other words, Leaman and Flanagan (2013) measured meta-language (or pattern language) used when describing research-based strategies. The current study analysed pattern language descriptions of research-based strategies via the SOLO taxonomy. In the current study, pattern language was not a standalone aspect of measuring a course’s effectiveness on PSTs’ ability, rather, one of a constellation of factors. PCK was seen as a prerequisite, and pattern language was taken into account along with attitudes and feelings of self-efficacy of PSTs, in terms of factors that would impact on practice.

Results found here are similar to the results found for PCK in that the embedded design participants scored higher results towards the end of the course than those in the applied experience. It is important to consider this similarity as mastery level
knowledge and understanding of PCK would inform the use of pattern language used in collaborative interactions. Although not studied here, the more extensive use of collaborative learning by the embedded design participants may have enhanced learning at a deeper level of understanding and retention of that learning more than other means of study.

The findings described here support the role of embedded design and applied experience in increasing the likelihood that early career teachers can contribute professional knowledge about inclusion with colleagues with whom they engage. The findings also signal the importance of ensuring that learning communities include professional language of the inclusive education field as a term of reference for effective collaboration. Hughes (2009) concurs that teacher preparation activities such as “dialogues with colleagues” and “cooperative group discourse” are effective strategies in enhancing field experiences and they allow PSTs to relate their experiences to theory (p. 255). Even though the highest levels of sophistication within the SOLO framework (Biggs, 1982) were not achieved, and the differences were not statistically significant between the cohorts, the results of this study provide important formative or emergent information about pattern language acquisition that can inform the way PL development is addressed in an inclusive education course.

After establishing support for the theoretical model of embedded design by demonstrating the positive impact on student learning of PCK, the second step was to see whether the learning translated into the development and sustainability of professional pattern language required for the application of learning in practice. Qualified support can be established for the impact of embedded design on PST knowledge and skills using professional language. On all occasions, embedded design scored higher than applied experience, even though some differences between the conditions were significant and others were not. The next step was to determine the differential effect of the embedded design and applied experience conditions on the self-efficacy of PSTs.

5.3 Self-Efficacy (SE)

Self-efficacy is said to have potential to be a strong predictor of whether teachers will implement and persist with effective pedagogies when working with students who have learning difficulties (Holzberger et al., 2013; Woodcock et al., 2012). There have
been mixed findings about the impact of self-efficacy (e.g. Hastings, 2012; Holzberger et al., 2013; Lancaster & Bain, 2007). This study examined how self-efficacy fits into a broader picture described in the theoretical model; and whether any changes in self-efficacy covaried with the independent variable, and in relation to the other variables studied. These were possible sources of self-efficacy based on Bandura’s theory (Lastrapes and Negishi, 2012; Palmer, 2006; Swackhamer, Koellner, Basile, & Kimbrough, 2009).

5.3.1 The research questions.

Does the application of the embedded design principle to the design of a PST inclusive education course results in improvements in self-efficacy?

Does the application of the embedded design principle covary with improvements in self-efficacy?

5.3.2 Summary of research findings.

In study one. The null hypothesis H_{01} c was supported as there were no statistically significant differences in the levels of self-efficacy between the PSTs who experienced application of the embedded design principle to the theoretical design of an inclusive education course as opposed to those who participated in an applied experience condition. The null hypothesis H_{02}e was also supported in study two where descriptive statistics showed there was no difference between self-efficacy between the PSTs who have experienced embedded course design when implementing ET compared to those who experienced the applied experience course design.

For study one, self-efficacy was assessed at pre- and post-test (year 2 session 3) of the PST program at the same time as PCK pre- and post-scores were taken. It was assessed again using matched samples of participants to compare post-test (session 3) occasion with post-prac (session 5) of the PST program for each condition using the SEIPD scale (Self-Efficacy toward Interactions with People with Disabilities) (Hickson, 1995).

The changes in self-efficacy scores on the SEIPD across time were statistically significant for both groups, that is, self-efficacy improved for both treatment conditions from the first to second testing occasion in study one. The embedded design condition began with a mean score of 80.39 at occasion one and increased to a
mean of 94.06 for occasion two. The matched samples went from slightly higher mean of 97.68 (session 3) and remained almost the same at occasion three 12 months later with a mean of 97.11 (session 5). The applied experience condition began with a lower mean at occasion one of 76.65, increased to a mean of 94.65 at occasion two. The matched samples for this condition went from 97.9 (session 3) up to 101.1 (session 6) showing that those in the applied experience condition improved their level of self-efficacy by the larger amount even though not of a magnitude to be practically significant.

5.3.3 Discussion of the results.
Results from this study concur with many others in the literature in that self-efficacy can be increased following the completion of a course of study about inclusion at undergraduate level (Bruder et al., 2013; Chiner & Cardona, 2013; Cologon, 2012; Eriks-Brophy & Whittingham, 2013; Gedžune & Gedžune, 2013; Haq & Mundia, 2012; Malak, 2013). All of these studies were empirical and directly measured the construct of self-efficacy related to inclusion. The dependent variables in this study incorporated direct measures of self-efficacy in addition to other mediating variables such as pedagogical content knowledge and pattern language.

Various course structures were described in the literature and then attributed to changes in self-efficacy, including the use of action research or the impact of different types of field-based placements. What was lacking in these aforementioned studies was a clear description of what actually constituted the course structure in order to have a positive impact on self-efficacy, and measurements of other mediating variables apart from self-efficacy scores.

By comparison, the elements of the embedded design course were described in detail and as a result can be linked to Bandura’s sources of self-efficacy (1989). These sources included: enactive mastery experiences, vicarious experiences, social persuasion, and physiological indexes that were incorporated into the course design in the following ways in the embedded design condition. Enactive Mastery was expected by way of the presentations to peers given during structured group class time. Vicarious experiences occurred through PST presentations and also in the collaborative groups where PSTs were enabled to estimate their capabilities in comparison to others as they worked together. Verbal persuasion was experienced
through feedback given by peers as well as by the instructor during workshop activities; and physiological indexes were often commented upon prior to quizzes, during presentations in front of their peers, and in the collaborative process of lesson design preparation. All of these sources were incorporated into the design of the course through the principle of embedded design, which was woven throughout all aspects of the course. Apart from vicarious experiences, these sources were not evident in the applied experience condition. Details of the course designs were provided in Chapter 3.

The nature and characteristics of the embedded design course also captured the additional sources of self-efficacy that had been identified by Palmer (2006) in his science classes, and Swackhamer et al. (2009) in relation to science and mathematics pedagogical content. For example, based on the description of mediating variables given by Palmer (2006), the embedded design ensured ‘content mastery’ which involved PSTs ‘knowing and understanding’ as opposed to just doing. This was enhanced by the weekly quizzes and the presentations.

Content mastery was expected by way of the differing assessment items required with knowledge being expanded across the quizzes and presentations, then finally culminating in lesson designs that incorporated the pedagogy of choice as well as iterations of differentiation that may be required in the classroom. The content was presented in a way that mastery was ensured before moving on to the next assessment item. Pedagogical Mastery was the key focus of content covered in the course and was scaffolded formatively via differentiated lesson designs. PSTs in both conditions also witnessed expert models during class, with the instructor modelling pedagogies, followed by peers modelling them as well. Palmer (2006) found ‘cognitive pedagogical mastery’ to be the most reported sources of efficacy when analysing his students’ reflections of their learning. In this study, both conditions were expected to reach mastery levels of PCK. Swackhamer et al. (2009) found that greater levels of PCK went hand in hand with higher levels of self-efficacy. For example, studies have shown that teachers with high levels of self-efficacy persist longer with students that struggle, recognise student errors, and attempt more efficacious strategies (Swackhamer et al., 2009).
This study concurs with the findings by Palmer (2006) and Swackhamer et al. (2009) as levels of self-efficacy did improve for both groups. Lastrapes and Negishi (2012) were able to extrapolate percentages of perceived importance that PSTs attached to the different sources of self-efficacy: 70% to their own mastery experiences, 20% to vicarious experiences, and the remaining 10% from verbal persuasion. All of these elements were incorporated in the current course being investigated through the use of embedded design principles, even though it was a classroom-based course rather than a field experience. These percentages do not help to explain how PSTs in the embedded design condition scored higher in the PCK quizzes (and pre/post-tests) than those in the applied experience condition yet recorded lower overall levels of self-efficacy. The results indicate support for the model, although unlike other elements of PCK and PL, the support was not at a level greater than the comparison applied experience condition.

Despite the presence of efficacy related features, the embedded design condition scored lower at occasion two and three compared to those in the applied experience; even though they had made higher gains in mediating variables of PCK and sophistication of language. Perhaps this lower score is a reflection of their more sophisticated understanding of the process of inclusion and what is involved. This is an important target for future research.

This study showed gains in self-efficacy for all PSTs involved and supported existing research indicating that a course experience in inclusion exerts a positive impact on PSTs’ assessment of their own confidence and capabilities relating to teaching students with additional needs (Holzberger et al., 2013; Woodcock et al., 2012). The results reaffirm the findings from the 2007 study by Lancaster and Bain who showed there were no statistically significant differences in levels of PST self-efficacy between three quite distinct approaches described in their study including embedded design. This study did not demonstrate statistically significant differences between the conditions under investigation at any occasion. Rather, these findings reinforced conjecture about the contribution of a direct applied experience to a pre-service preparation course and the influence on levels of self-efficacy. Given that mastery learning of PCK and better results for pattern language were associated with the embedded design condition, the results of this study add to the conjecture over the kind of experiences that are more or
less effective in improving self-efficacy as it relates to working with students with different learning needs.

The findings stand in contrast to findings from Yilmaz (2011), who determined that the more proficient teachers perceived themselves to be with PCK (for example, about explicit teaching and cooperative learning), the more efficacious they felt in implementation of instructional pedagogies, and the higher the levels of their self-efficacy. A teacher’s sense of efficacy directly influences the kind of environment they create for their students ensure learning (Yilmaz, 2011). The relationship between self-efficacy and environments created by teachers may be more complex than thought and worthy of further exploration.

After establishing support for the theoretical model by demonstrating the positive impact on student learning of PCK, the development and sustainability of professional pattern language requires the application of learning in practice. There was improvement in self-efficacy for the embedded design condition, but not over and above that associated with the applied experience condition, despite the embedded design possessing more features that were sources of efficacy. The variable of self-efficacy, however, is not seen as a standalone influence of PST success with inclusive practices, rather, it is one of the integrated factors that will work together and translate into better practices in-situ in an applied setting.

It is also possible that self-efficacy is not as significant as the influence of PCK or the use of sophisticated language when it comes to translation of these skills and knowledges into actual practice. Few self-efficacy studies have had the opportunity to examine ratings using a criterion variable of actual practice. What this study shows in one instance is that the construct may be influenced by mediating factors other than those of mastery experiences and vicarious learning opportunities, and this warrants further investigation. It is possible for a person to report perceived efficacy at a level that may not covary with their actual practice. Results indicate a need for more self-efficacy research where practice is a criterion.

We should not underestimate the value of applied experience which underpins assumptions about practicum and other professional experiences as key components of pre-service preparation. The following section will discuss whether the skills and
attitudes investigated in study one translated into practice for study two; in other words, to determine the differential impact of course design on the design integrity, implementation integrity, and self-reflection integrity of PSTs’ inclusive classroom practices.

5.4 Design, Implementation, and Self-Reflection Integrity of Inclusive Classroom Pedagogies

Design, implementation, and self-reflection integrity are among the most significant variables indicated in Figure 5.1. They refer to the ability of PSTs to design and deliver instruction using the research-based pedagogies they had been studying in classes at university 12 months earlier and then to reflect on the design and delivery after the fact. The research design illustrated in Figure 5.1 is the final step in determining differential impacts of course design on what PSTs know, how they can discuss that knowledge in sophisticated ways, and how they design a lesson plan and then translate these skills, knowledges and attitudes into practice. Implementation has three dimensions to consider and will be discussed next: that of design, actual delivery of the lesson, and then reflection by PSTs.

5.4.1 The research questions.

Does the application of the embedded design principle have an impact on the fidelity that PSTs use to design lessons that include the research-based characteristics of inclusive pedagogy?

Does the application of the embedded design principle have an impact on the fidelity that PSTs use to implement lessons that include the research-based characteristics of inclusive pedagogy?

Does the application of the embedded design principle have an impact on the fidelity that PSTs use to self-reflect on lessons that include the research-based characteristics of inclusive pedagogy?

5.4.2 Summary of research findings.

There were three components of inclusive education pedagogy integrity used in this study: (i) the PSTs’ ability to design the pedagogy to be implemented and demonstrate that design through a lesson plan, (ii) the actual implementation of that pedagogy, and finally (iii) the self-reflection about the design and practice. The design of the practice
was assessed using document analysis of the PST lesson plan. The actual implementation was assessed using researcher observation and then PSTs’ self-evaluation occurred following delivery of the lesson. An integrity protocol was devised by the researcher to reflect the characteristics of explicit teaching lessons as specified during university coursework. The use of these three different data sources assisted with reliability of data collection as the same protocol was used in each instance.

Inferential analysis was not applied to implementation integrity due to the small numbers of participants (the embedded design involved 21 participants and the applied experience involved 10 participants). While a number of procedures can be applied to address unequal group sizes, as had occurred in study 1, it was decided that simple descriptive analysis of the data derived from a criterion-based measure would most adequately represent the findings in study 2.

Document analysis involved the researcher scoring the written lesson plan provided by the PST using the same protocol employed for direct observation. The researcher scored the lesson designs noting the presence or absence of key features in the lesson. Researcher observations involved the researcher observing the lesson designated by the PST and noting observed characteristics on the same protocol, in this instance it was used for a direct observation of the lesson. The self-assessment of design and implementation also used the protocol and involved the PST scoring the same lesson following lesson delivery.

The mean scores for document analysis of implementation fidelity in the embedded design condition (M =15.36) were 9.46/24 points higher than for the applied experience condition (M = 5.9). The PSTs in the embedded design condition were more often able to include essential characteristics of explicit teaching in written lesson documentation, than those of the applied experience condition.

For classroom observation of the lesson implementation, the mean scores as rated by the researcher, for PSTs in the embedded design condition (M = 16.57, out of a possible 24) were higher than for those in the applied experience condition (M = 9.00, out of a possible 24) with a difference in mean scores of 7.57 (32%) in favour of the embedded design condition. Neither of the conditions reached mastery levels here. However, achieving
approximately 70% was an encouraging finding given the 12-month time period between the course content and the practicum. While the embedded design teaching approach employed was pivotal to inclusive practice, it was not reiterated or developed through subsequent course experience. The closeness in scores between document analysis and researcher observation could be expected and attributed to the fact that the researcher scored both activities. Reliability scoring was incorporated to offset any bias. It would also be expected that if characteristics were absent from the plan they would not be expected to occur in practice. This indicates that if the design characteristic is not present in the lesson plan, it is likely that the characteristic will not be observed in the observation of that lesson.

Results for PSTs’ self-evaluation of implementation fidelity found the mean score for the embedded design condition \( (M = 17.05) \) was 4.15/24 points higher than for the applied experience condition \( (M = 12.9) \). The PSTs in the embedded design condition perceived themselves to be incorporating essential characteristics of explicit teaching after implementing a lesson in a school setting at a higher level than those in the applied experience condition. Implementation of the lesson could only be as good as the design of that lesson and the inclusion of key research-based features.

These results suggest PSTs in the embedded design condition designed and implemented explicit teaching lessons with a higher degree of implementation fidelity than those PSTs in the applied experience condition. This finding is based on data analysed from researcher observation, self-report and lesson plan document analysis.

The following section will elaborate on the effectiveness of the courses in translating requisite skills into practice.

5.4.3 Discussion of the research findings.

Implementation integrity of inclusive education pedagogies is of utmost importance if the research-based characteristics of such pedagogies are to translate into practice and affect learner outcomes (Cook & Tankersley, 2012; Klingner et al., 2004; Kretlow & Helf, 2013). Kretlow and Helf (2013) suggest that “when evidence-based … instruction is implemented with fidelity, instruction can be eliminated as a reason for students not making adequate progress” (p. 168). A sound understanding of
pedagogical content knowledge about research-based pedagogies will assist teachers in translating them into practice.

When determining the effectiveness of translating requisite inclusive education pedagogies into practice, comparisons were enabled between the embedded design course and the applied experience using descriptive statistics such as means and standard deviations. In spite of major studies in the past calling for a more empirical evaluation of PST programs as the means to close this research to practice gap (Cochran-Smith & Ziechner, 2005), no other studies were located in the time period since that involved this quantitative evaluation.

### 5.4.3.1 Design fidelity.

The hypothesis from study two, H2a, was supported and results indicated through descriptive statistics showed higher levels of design integrity of research based pedagogies for PSTs who have experienced embedded course design when implementing ET compared to those who experienced the applied experience course design.

While participating in a practicum experience, all PSTs chose to deliver lessons using explicit teaching pedagogy for the observations. The scoring conditions for design fidelity through document analysis indicated that the task of developing a criterion scored the lowest of the three elements in the learner outcome section of the protocol. The percent correct for the three elements of the learner outcome in the lesson design were comparable to researcher observation scores and self-assessment scores for the embedded design condition. In the applied experience condition, however, there was zero evidence of the learner outcome characteristics. Participants in the applied experience condition did not attempt any elements of the learner outcome. Instead of noting specific condition, actions and criteria in their documentation, they included outcomes taken directly from the syllabus.

The characteristic of modelling within explicit teaching was achieved at the level of 70% (AE=7) correct for the applied experience condition, and lower for the embedded design condition (ED=8, 61.5%). PSTs in the applied experience condition scored 10% (AE=1) for ‘characteristics of guided practice broken down into logical steps’, and the same for making sure that ‘guided practice related to the model’ (AE=1).
PSTs in the *embedded design* condition achieved 75% correct (ED=8) for ‘guided practice broken down into logical steps’ and 58.3 % (ED=8) correct for making sure the ‘guided practice related to the model’. Relatedly, Hattie (2009) notes that the process of working through an example is the most critical part of explicit teaching and with low figures like these, the step of *modelling* is not being used to maximum advantage.

The differences between cohorts were also evident for the *differentiation* characteristics of explicit teaching. Participants in the *embedded design* condition had written the characteristics of *differentiated content, product and criteria* in their lesson plan documentation. Participants in the *applied experience* did not include any differentiation possibilities in their written lesson planning. Almost 85% (ED=20) of those in the *embedded design* group were able to differentiate the product, action or criteria required to cater for different learning needs. As this is an essential skill in an inclusive classroom, the ability to differentiate content is of utmost importance.

This format of data collection (document analysis) highlighted pedagogical characteristics in PST planning and design of the lessons they were about to teach. While the design for both groups was incomplete, participants from the *embedded design* condition scored higher results and included more of the essential characteristics than the *applied experience* condition. Areas of particular strength for the *embedded design* condition were correct design of the learner outcome, guided practice steps and independent practice. Based on these findings, direct researcher observation of practice could be expected to illustrate the same omissions translated into practice in the classroom.

### 5.4.3.2 Implementation fidelity

The research hypotheses for study two, H2 b, was supported and indicated that descriptive statistics for classroom implementation showed higher levels of Implementation integrity for PSTs who have experienced *embedded* course design when implementing ET compared to those who experienced the *applied* experience course.

The fact that important characteristics were not included in a lesson plan might explain why the same characteristics were not evident in the actual delivery of the lesson.
Direct researcher observation indicated that the first three scale items from the observation protocol for explicit teaching (condition, action and criterion) constituted the core items associated with writing a learner outcome for an explicit teaching lesson. In all instances, participants in the embedded design condition implemented these characteristics at a higher level of fidelity than those in the applied experience condition. The strengths of the PSTs lay with the characteristics of: providing an attention grabber (that is something to gain the attention of students), identifying key words, and linking the lesson to prior learning with both conditions scoring 80% and above. Modelling was an issue for both with scores of 60% (AE=6) - 66.7% (ED=14) indicating the teacher model was only partially evident. Just as for document analysis, the embedded design condition scored higher for guided practice being broken down into steps 71.45% (ED=15) compared to the applied experience condition 20% (AE=2); and also for independent practice where the embedded design condition scored 81% (ED=17) compared to the applied experience condition 30% (AE=3).

The overall weaknesses in implementation were seen in linking the guided and independent practice to the actual model that had been given. The characteristics that were observed by the researcher coincided with misalignment with the quality of the outcomes. It is important for all features and characteristics of explicit teaching to be present. If connection between modelling and guided or independent practice of the model is poor then it is hard for the guided or independent practice to be effective.

These implementation results align with the PCK results achieved when participants were studying the university course 12 months earlier. At post test for study one, participants in the embedded design condition scored mastery level of PCK scores that were statistically significantly higher than participants in the applied experience condition. Again, results demonstrate that these characteristics need to be understood to a level of mastery before they can be generalised by PSTs to actual design and implementation in practice.

Differentiation between the PST groups was also evident for the components of modelling, use of guided practice, and independent practice. Even though embedded design scored higher than the applied experience condition, neither group reached mastery in terms of linking their model directly back to the action in their learner
outcome. Here, PCK of a particular characteristic is one level of PST ability, while the
demonstration of that characteristic appears to be more difficult.

5.4.3.3 Self-evaluation fidelity.
The research hypotheses for study two, H2c, was supported by descriptive statistics for
classroom implementation that showed higher levels of self-awareness about lesson
design and implementation for PSTs who have experienced *embedded* course design
when implementing ET compared to those who experienced the *applied* experience
course design.

Evaluations of lessons based on self-assessment by the PSTs indicated whether there
was an awareness of strengths or faults in their planning and design. The PSTs’ *self-
evaluation scoring* revealed that PSTs from both conditions scored themselves higher than
the researcher on half of the scale items, which indicates they were not aware of the errors
in implementation they were making. The scores from the *embedded design* cohort were
closest to those of the researcher.

One of the largest differences between researcher observation and the self-scoring was on
the scale item of *criterion* in developing an outcome for the lesson. Clearly the PSTs
thought they were addressing the need to have criteria explicitly embedded in their lesson
plan, while the researcher did not observe this to be the case. The perceptions of the PSTs
did not line up with actual performance.

The differences here between PST self-evaluation and researcher observations should be
considered in light of the scores found in the self-efficacy data. It is clear for both groups
that their perceptions about their implementation integrity did not line up with the
observations, and that this misperception was greatest for the *applied experience* cohort.
The PSTs may have a false sense of their own mastery of these elements, especially those
in the *applied experience* design, which was reflected in higher self-efficacy levels. Given
that self-efficacy is a construct measured by self-report and not necessarily validated in
actual practice, it is possible that inaccurate perceptions of expertise could be reflected
in reported self-efficacy levels by PSTs. The self-report of lesson implementation
followed actual implementation of the lesson and perhaps the natural feedback from school
students during the delivery of a lesson served to lower PST perceptions about their self-
efficacy.
When investigating the research-to-practice gap and why PSTs do not always translate their university learning into practices, the existing literature for PSTs focused on implementation integrity in two main formats: the use of simulated settings and of in-situ implementation. This thesis included data from both simulated settings (in terms of a course experience about pedagogical content knowledge and pattern language); as well as data collected in practicum settings (in-situ). The simulated settings provided the arena for modelling and guided practice of pedagogies during tutorial times. The in-situ settings provided the arena for independent practice of those pedagogies during the PSTs’ practicum; and this applied to both conditions. The implementation integrity analysis was designed to see whether PSTs could apply what they learned. No other studies were located that investigated these two scenarios in the same manner; comparisons between these formats will now be discussed.

5.5 Simulated Settings

Most other studies set up simulated scenarios, while this current study was a real world application in the course and in the practicum. Empirically evaluated inclusive education skills in simulated settings were found in studies by: Florian and Rouse (2009); Fox-Turnbull and Snape (2011); and Leaman and Flanagan (2013). This study was also empirical and utilised direct measures to collect data (in terms of PCK and pattern language). The current study investigated similar PST content areas to those found in Florian and Rouse (2009). For example: Florian and Rouse (2009) asked their PSTs to make direct observations of lectures provided by staff to see if PSTs could code their observations about inclusive pedagogy in relation to Schulman’s three apprenticeship model including the head, the hand and the heart (in Florian & Rouse, 2009). These content areas within Schulman’s model can be related to content covered in the current study as follows: the head (knowledge) is related to PCK and knowledge about professional pattern language; the hand (skill and doing) is related to observations of PSTs in schools on their practicum; and the heart (attitudes and beliefs about inclusion) relates to self-efficacy about inclusion. The study by Florian and Rouse (2009) analysed whether PSTs could recognise the features of good practice, not about their knowledge and skill with those same features in practice.

These elements were all directly measured in the current study as they were demonstrated by the PSTs. The current study measured skills and attitudes directly
(relating to the head, hands and heart from Schulman) while in the simulated settings of tutorials. The direct measures for Florian and Rouse (2009) focused on the categorisation of practice to determine whether any praxis existed with Schulman’s framework. PSTs in the current study were assessed directly on the development of their skills, knowledge and attitudes as a result of the course completion (and translation of those skills, knowledge and attitudes into practice settings). If the PSTs had described characteristics while watching a lecture, this would have been significantly different to actually conducting the lesson themselves.

5.6. In-situ Settings

In a similar way to other studies, the current study also analysed PST skills while the PSTs were on practicum (in practice) (for example: Klein et al., 2013; Lamb, et al., 2013; Lynch & Smith, 2012; Moran, 2014; Sela & Harel, 2012). These studies failed to measure actual changes of the PSTs’ skills, knowledge and practices. Rather, they focused on issues other than those indicators in the literature that relate to successful inclusion (Lynch & Smith, 2012; Moran, 2014; Sela & Harel, 2012). These authors did not use any measures to determine whether a PST course actually worked, and they did not determine if there was actually any change in PST behaviour as a result of knowledge and skill development in real settings. The recognised inclusive education indicator, as reported in their abstracts, was often measured indirectly through Sela and Harel (2012) measured reflections following the use of an action research project whilst PSTs were on practicum. Data revealed feelings about using action research as a process rather than observing the effects on PST knowledge and skills. Lamb et al. measured reflective practice based on observations made by peer reviewers. Lynch and Smith (2012) also relied on PST perceptions about their practicum experiences.

PST perceptions of their practicum experiences reported by Lynch and Smith (2012) resonate with results found in the current study. In the analysis of the self-efficacy (SEIPD) results, reflections and perceptions expressed provided a less accurate picture of PST ability when observing translation of those reflections and perceptions into practice. This is an important similarity between the results noted in the Lynch and Smith study, and results found in self-reports by PSTs in the current study, that the self-reports PSTs were largely inaccurate.
Self-reports in the current study about PST skills in-situ were in direct contrast to the results found following administration of the SEIPD. Study 2 results of SEIPD scores for embedded design remained stable with similar means between occasion 2, session 3 (at the conclusion of the course) and occasion 3, session 5 (12 months later on practicum), while those in the applied experience condition increased their mean SEIPD scores at the final occasion, yet their scores on the implementation protocol were lower. The results from the implementation data collection stand in contrast to the self-efficacy scores where PSTs in the applied experience condition scored higher. The self-report of lesson implementation followed actual implementation of the lesson. The natural feedback from school students during the delivery of a lesson may have served to lower PST perceptions.

Regardless of whether other authors used simulated or real settings, they did not measure actual changes of the PSTs’ skills, knowledge and practices. Rather they focused on indirect measures of issues other than those indicators in the literature that relate to successful inclusion (Lynch & Smith, 2012; Moran, 2014; Sela & Harel, 2012). The current study directly measured recognised indicators of high quality implementation of pedagogies whilst PSTs were completing their practicum placement. The research methodologies others incorporated in their articles (i.e., case study, action research, and reflective practice) were not used to directly measure relative inclusive skills and knowledge required by PSTs (Klein et al., 2013).

Not only did the current study incorporate direct measures of PST skills and attitudes both in simulated settings as well as in practice, the current study was also able to differentiate between course designs (those of embedded design and applied experience) by way of comparisons on the indicators of inclusion. These findings constitute several steps in building support for the conceptual principle of embedded design by demonstrating the positive impact of the embedded design course experience on PST learning about PCK, the sophisticated use of pattern language, and the generalisation of these to practice.

To date, the variables that contribute to successful inclusion have been discussed in the preceding sections of chapter 5. The discussion now turns to the theoretical drivers behind course designs that assist with translation of required practices into the field. The links between theory and actual practice will now be discussed.
5.7 Links between a Theoretically Designed Course and PST Practice

The current literature provides evidence of a serious research to practice gap regarding PSTs’ skills, knowledge and attitudes (Antil et al., 1998; Grima-Farrell, 2012; Grima-Farrell et al., 2011; Hennessey & Dionigi, 2013; Jenkins et al., 2003). This is a disconcerting finding given the benefits of these practices are based upon research studies where their procedural features were implemented with high degrees of integrity (Klingner et al., 2004), and the scant research that directly measures PST abilities.

In this study, the embedded design condition derived from a theoretical model of self-organization (Bain, 2007) was compared to an applied experience in the areas of PCK, pattern language, self-efficacy, lesson design, lesson delivery and reflection. The study made comparisons between the embedded design condition and the applied experience condition to determine differential effects related to the independent variable. Unlike Jin et al. (2013) who found in their study that the curriculum could be related to theoretical principles of the Trans Theoretical model, the results here empirically determined differential effects between cohorts in favour of the embedded design condition over the applied experience condition. The achievement distribution suggested that the application of the embedded design principle improved understanding of inclusive practice over the applied experience approach. The additional 11 hours of applied experience in real settings did not result in increased achievement levels over those recorded in the embedded design condition.

PSTs from the embedded design cohort were able to display higher levels of implementation characteristics when working in the classroom even though they displayed lower levels of self-efficacy. Overall, the findings lend support to the statements from Jamil (2012) and Bandura (1989, 1997) that self-efficacy beliefs predict the choice of task, effort expended, and persistence shown--and ultimately the level of student success achieved (Bandura, 1989, 1997). This study has linked mediating variables of self-efficacy (as independent variables) and measured them.

An obvious potential for future study would be to use these results to inform the broader conceptual framework for a PST education program where theoretical course design, course content delivered through well-structured learning communities, as well
as an *applied experience*, could be integrated. The comparison conditions could be used in combination with *embedded design*, although this was not studied here. This resonates with findings from Kim (2011), who suggest that undergraduate PSTs rank highly the provision of an environment that scaffolds their learning and promotes effective interaction amongst participants. The scaffolded approach inherent to *embedded design* in combination with *applied experience* would allow for such a scaffolded learning environment. The combination of collaborative problem solving, a shared professional lexicon, and field experience could shape up to be potentially powerful learning tools for PSTs that could be investigated in future.

As with Fox-Turnbull and Snape (2011), the current course used collaborative groups to enhance PST learning during their *simulated settings* in tutorial sessions. Group work was incorporated to embed the design through collaborative teams while PSTs were working on lesson plan drafts and giving each other feedback about required characteristics of pedagogical design. The actual collaboration process was not a dependent variable or focus of analysis, but rather the means to construct deeper levels of knowledge of PSTs. Like Florian and Rouse (2009), Fox-Turnbull and Snape (2011) and Dooly and Sadler (2013) did not measure PST skills or knowledge, but rather they measured PST ability to relate their method of learning to constructivist theories and categorisation of practice based on the chosen theory.

The findings for the first three research areas constitute initial steps in building support for the conceptual model described in Figure 5.1. Results demonstrate a positive and differential impact of a course experience on PST learning about inclusive pedagogies, use and sophistication of pattern language, and levels of self-efficacy. These findings support the idea that carefully monitored course design can have the capacity to build mastery level knowledge and reduce the variance in performance on assessments of that knowledge.

This study also demonstrated an initial contribution of the *embedded design* principle where results illustrated the differential effects of skills and generalisation of these PST skills into actual practice. Overall, the findings provide support for a course utilising the theoretical principle of *embedded design* in order to deliver these findings. The fact that none of the groups achieved mastery levels of performance whilst on
practicum indicates a need for more deliberative preparation if PST students are to be able to demonstrate mastery practice in applied settings.

As well as these positive results, limitations of the study must be considered. The following section will discuss the general limitations of the study, and finally, suggest directions for future research.

5.8 Limitations of the Study

This section identifies and discusses the limitations of this research. These include limitations due to: design; and the reliability of the dependent measures.

5.8.1 Design.

The first limitation is concerned with the quasi-experimental design of the study given the number of threats to validity identified and discussed in Chapter 3. The major threats to internal or external validity were identified as: selection bias, as well as additive and interactive effects. The threat of history was the biggest single threat to this study because of the delays to the practicum (Shaughnessy et al., 2009).

Other elements of validity were considered in the design of the research. This study’s generalisability is clearly limited by its focus on just one university course across two settings, an available population of students, and the quasi-experimental nature of the design. The quasi-experimental design was the most appropriate to use as the random assignment of participants to control and treatment groups was not possible due to the settings from which the participants were drawn. PST participants were already in class groups at different settings. When random selection is not an option, as in the current study, groups can be selected that are as similar as possible and a pre-test of dependent measures administered to check for initial equivalence.

To reduce the threat of selection biases in study one of this research, pre-occasion data was collected on all dependent variables of interest, as well as demographic information, in order to check for nuances between the groups. The results at pre-test level for content mastery, self-efficacy, and use of professional language indicated no significant differences and were reported in full in the results section (Chapter 4). Pre-intervention data was also used to minimise the threat to interactive effects on validity. Any results that indicated extreme values in the pre- or post-data occasions were
excluded from the data set to minimise this threat. Thus, evidence strongly suggested the participant groups were close to equivalent at the time of the pre-test occasions. Sufficient detail about the participants and the dependant variables is provided throughout the thesis to allow for replication which would enhance external validity.

5.8.1.1 Statistical conclusion validity.
Reliability of treatment implementation and extraneous variance in the experimentation setting was the only issue to be considered a major concern based on discussions in Chapter 3. The variations in the way the treatment condition (the independent variable of course design) is implemented can inflate error variance and may decrease the chance of obtaining true differences (Burns, 2000; Fraenkel & Wallen, 2006). Within the context of this study, internal validity could have been threatened by groups being exposed to different teachers. Two of the embedded design tutorial groups were conducted by the researcher; two further tutorials were conducted by a colleague as were the tutorials on the applied experience campus. The lectures were shared between colleagues. The implication here is that consideration needs to be made of the way the course may have been taught in different ways by three different personnel. This threat was reduced by trying to standardise the treatment across all tutorial groups. The applied experience cohort was not utilising embedded design in delivery of the course (rather an applied experience design) so close attention was required to report exactly the way the course was delivered at each setting. Procedures were put into place to document treatment implementation fidelity across settings. Data collection for the course implementation fidelity may be found in Appendix 1: “Course implementation data collection tool across campus delivery”. However, completing these forms did not guarantee that differences in delivery were accurately identified.

A further limitation of the current study is the small sample size. Even though it was possible to run statistical analysis with the data from study one, the inequity in the group sizes for the different conditions presents a reliability issue. This is even more critical with the observational data from study two, whilst the PSTs were on practicum. Inter-rater reliability went part way to minimising the risks here and 80% agreement was achieved.
It is often the case that the independent variable may be confounded by extraneous variables, in the categories of participant or environmental variables, which make it difficult to determine the unique effect of each. Randomisation of treatments or participants is the best single way to control for many extraneous variables; however, as noted above, it is not always possible (Burns, 2000; Lankshear & Knobel, 2004; Shaughnessy et al., 2009). The second main area for potential limitations concerns the reliability of the dependent measures that were incorporated for data collection.

5.8.1.2 History.
History was the most significant threat to the current study. Initially, the practicum data collection was to occur four weeks after completion of the university course on inclusion. The delay in ethics approval meant that this data collection did not occur until 12 months later and in this time interval there may have been many outside influences that could have affected PST results. The unintentional time lapse between data gathering points may have had an impact on results and thus it is important not to attribute results to intervention alone. These outside influences remain unknown and out of the realm of control of the researcher.

5.8.2 Reliability of the dependent measures.
The Self-Efficacy toward Future Interactions with People with Disabilities Scale (SEIPD) (Hickson, 1995) was a standardised instrument employed during the study. The SEIPD was selected for this study as at the time of data collection, no other suitable scale had been developed that targeted inclusive education and PST self-efficacy.

The reliability of the SEIPD was originally determined using test-retest and alpha coefficients employing a sample of 180 PSTs and nurses (Gay & Airasian, 2003). A mean alpha co-efficient of .87 was reported for the SEIPD, while test-retest reliability produced a reliability coefficient of .8 over a four-week interval (Hickson, 1995). Factorial validity was established using principal component analysis, and both orthogonal and oblique rotations gave identical results with only one factor extracted, indicating that items within the scale were measuring the same construct and accounting for an average of 55.1% of the variance (Hickson, 1995).
Construct validity refers to the constructs that are actually being measured by the test (Burns, 2000; Cohen et al., 2007; Gay & Airasian, 2003). Statistically significant correlations (.32) were found confirming the mediating relationship of attitude on self-efficacy. The current use of the instrument was altered to include the word ‘student’ more frequently so it would hold more ‘face validity’ for the PSTs. For example: item 1 originally read “I feel confident in my ability to work with individuals with disabilities” but was changed to “I feel confident in my ability to teach students with disabilities”. Making even seemingly small changes to an instrument that has already been standardised may have had an impact on results.

As with studies by Block (2013) and Hannman (2013), this SEIPD instrument was also subjected to quantitative validation in a study by Zundans-Fraser and Lancaster (2012). The original SEIPD scale developed by Hickson in 1995 was employed in this study. At the time of Hickson’s validation, the scale consisted of only one factor. In the study by Zundans-Fraser and Lancaster (2012), factor analysis yielded two interpretable factors: personal teaching efficacy (PST) and skill level. The two factors that were noted in that 2012 study continue to strongly relate to Bandura’s (1989) notion that the construct of self-efficacy is mediated by multiple variables. The two factors also fit well with the variables of successful inclusion studied here and supported by results from Lee et al. (2011).

Even though self-report is a direct measure, caution must be exercised as it was only possible to report on pre and post differences based on the self-report of self-efficacy. Findings generated from self-report are difficult to generalise unless there are other data sources to support the findings. The utility of self-efficacy is best contextualised when reported in studies where actual practice can also be studied. PSTs were predisposed to the same questions at pre- and post-test occasions so they were aware of what was expected of them by the post-test occasion.

Limitations may be noted when analysing results from the instrument used to determine *pedagogical content knowledge* as it was a criterion-referenced test that involved a comparison against predetermined levels of performance rather than against other PSTs (Gay & Airasian, 2003). The PCK scores were derived from a set of 20 researcher-developed questions that framed the content to be covered
throughout the course. There is also the potential for a testing effect related to the use of the same criterion measure in pre- and post-test conditions.

There were a number of factors that could affect reliability (Cohen et al., 2007, pp. 159-163), and a range of moderation strategies were used both before and after the tests to enhance the reliability of researcher made tests including: group moderation of grades; post hoc adjustments to grades; defining marking criteria; providing exemplification; and holding group moderation of grades meetings. Agreement trials and practice marking for inter-scorer reliability were all undertaken immediately following each PCK test or quiz (see the Chapter 4 for details of inter-rater reliability obtained). Reliability of marking was achieved between researcher, colleagues and a research volunteer by jointly grading 20% of quizzes, randomly selected across tutorials. A decision table was constructed to facilitate coding and to ensure a high degree of consistency (Johnson & Christensen, 2008) to ensure that there was consistency between markers.

A potential limitation could be noted in the choice of content to be covered in the course. Members of the inclusive education team agreed on important operationalisation of the pedagogical construct. The definitions were confirmed through literature reviews of explicit teaching and cooperative learning research-based strategies used in the field. As a researcher constructed instrument, it is not appropriate to compare results to any other standardised formats available to judge construct validity.

The scoring of pattern language frequency involved similar limitations to PCK as required terms were researcher determined. All the same precautions were in place when analysing pattern language frequency as for PCK in order to minimise the limitations.

The sophistication of pattern language was analysed using the SOLO taxonomy developed by Biggs and Collis (1982) with high reliability ratings ranging from 0.79-.95. The current reliability was determined by using inter-rated reliability and achieved a score of more than 90%. This does not have the same empirical integrity as the factor analysis employed by Biggs and Collis (1982) or Boulten-Lewis (in 1994). It also needs to be mentioned that there is a possible floor effect where the higher levels of the
taxonomy were not achieved and possibly could not be expected to be achieved in just one course.

While the design of the present study does not permit direct causal inferences regarding the implementation of the embedded design principle, it is clear that an increase in PCK and pattern language covaried with the implementation of the principle expressed in each teaching cycle, and was more effective in developing these capacity building knowledges and skills when a full embedded design was employed (Bain, Lancaster, Zundans and Parkes, 2009).

A researcher developed observation checklist was used to investigate the translation of these capacity building skills into practice. The factors outlined by Zirpoli (2015) were taken into account to minimise the limitations of using observational data. However, the setting of the practicum may have influenced the detail included in the lesson plan as some PSTs were told they could progress to day book notes instead of writing up more detailed lesson plans. For the occasion of researcher observation, PSTs were asked to complete a full lesson plan, but this may not have been the most recent practice for them. The real concern is that the missing elements in the written documentation for the lesson were also missing in the delivery of the lesson. The core starting point for explicit teaching is the learner outcome. If any of these explicit teaching elements were incorrect or missing, it would be difficult for the PSTs to follow through with the rest of the lesson. The single element that caused most difficulty for both conditions was the criterion by which to judge student success. The mainstream curriculum documents supplied by the Department of Education (for example, the Board of Studies Teaching and Educational Standards from New South Wales) do not normally include specific criteria. All elements of learner outcome were, however, taught to mastery throughout the course for content knowledge results and this knowledge did not translate into the practice of PSTs. The potential for a washout effect or a history effect was evident as it was not possible to control what the PSTs experienced in that period. It was not controllable during the time lapse of 12 months between the completion of the internal course and practicum. A further consideration to take into account is that teacher skill would be on the continuum of a Proficient Teacher as per AITSL Standards (AITSL, 2015). That is, graduates are not expected to complete their programs and be proficient in all areas required. They would be working towards proficiency with ongoing learning once in the field.
The following section will address potential future directions.

5.9 Future Directions

The study has reinforced the need for more focus and research on the specific nature of the PST experience and the critical need for advocates to maintain a call for the empirical study of teacher education. For too long, teacher education institutions have employed inefficient means of evaluating their own practice. Even though this call for more accountability has been made for considerable time now, there has yet to be a convincing response from the field (AITSL, 2015; Florian, 2011; Ludlow et al., 2010).

This study represents a starting point and provides a direction for future research involving larger scale randomised trials. For example: the issues of random selection, quasi-experimental design, and means of enhancing the reliability and validity of data collection, could all be the focus of future research with the intent of being more accountable. With sufficient detail of the methods used here, repetition of the study or the addition of additional curriculum areas investigated in the same way could assist reliability. The PCK chosen here is not fixed or unalterable within the embedded design, and different content areas of PST programs could be designed based on the embedded design principle and tested for efficacy.

Repetition of the embedded design utilising differing variables of interest could include enhanced reliability through pre-test assessment of the dependent variables. This would ensure a level playing field where non-equivalent groups are used as a means of comparison as the random selection of participants is going to be a continuing issue within PST programs. Larger scale studies could also be subject to more sophisticated analysis and modelling to determine the relative contribution of the different factors.

The coherence of programs and the extent to which these programs build skills and knowledge in a coherent way needs further investigation to determine reasons for the washout issue and reduction in scores found in the current study.

Variations in the treatment implementation of chosen pedagogies could be addressed by the use of video recordings of the content being delivered by different staff. This could be achieved with the use of current technologies such as Adobe Connect online
meetings where sessions can be attended live from remote sites, and can be both recorded and analysed thoroughly and independently.

With due recognition of the limitations that have been previously noted, the findings in the study lend support to the potential of an educational design model, specifically the *embedded design* principle, to influence PST performance in an undergraduate course on inclusion. This is especially the case given the general paucity of data associated with the scalable influence of a design approach in teacher preparation. The findings also lend support to existing research on the *embedded design* principle that has shown significant gains in pattern language, pedagogical content knowledge and implementation integrity.

These findings could stimulate the continued examination of the role of theory-driven course design in teacher preparation programs. This includes an examination of the way in which the application of the *embedded design* principle across multiple courses in a program can contribute to higher levels of professional understanding and, as mentioned previously, the extent to which that covaries with change in classroom practice. As Sari, Celikoz and Secer (2009) suggest, perhaps additional inclusive education courses should be conducted for PSTs; or, as Forlin, Loreman, Sharma and Earle (2009) noted with their Canadian cohort, the inclusive education content might be embedded across a whole program rather than a standalone course within a program. Enactive mastery described by Palmer (2006) could be linked into professional placements following the conclusion of an inclusive education course. Further research is required before more definitive conclusions can be made.

The results here are indicative of the need for a careful examination of the connections between class coursework and the design of an *applied experience*. This examination could involve a combination of the *embedded design* and *applied experience* approaches described in this study in order to maximise the potential benefits of both. This work could then focus on the extent to which those experiences are mastery-based and the ways in which they connect to the future in-service role of PST education students. To investigate maintenance and generalisation of skills once graduation has occurred, a replication of the work such as that carried out by Grima-Farrell (2012) could be conducted.
There is no room to speculate about the long term effects of the embedded course design, and none were made. Nor are there any claims about the actual ability of PSTs to teach in inclusive settings following graduation. To make these types of claims a more longitudinal study is required, and in fact is desirable in order to continue the quest for accountability of course designs. This study has taken important steps on the trajectory from classroom and course experience to actual practice.

The theoretical drivers behind course design calls for a more thorough analysis in terms of the many variables that impact PSTs, for example, the embedded design condition was multicomponent as was the applied experience condition and a need exists for further research to determine which aspects of both approaches were driving the results achieved.

The mediating influences of self-efficacy also present a need to examine these more carefully. Given the inconclusive results from self-efficacy studies in the literature, more studies are needed where perceptions of efficacy are tested in practice. Mediating factors of self-efficacy are a link that has been suggested between mastery levels of PCK and self-efficacy. Further links could be investigated to see what is efficient in assisting PSTs to have a clearer or more accurate sense of their own abilities. The extant literature has produced mixed results beyond the usual “Yes, there was an increase between pre- and post-test data collection”.

A few authors have attempted to establish a mediating link between other variables, but very little has been attempted in terms of establishing differential links that might have a positive impact on levels of self-efficacy and the relationship to practice in real-world settings. Developing a more specific questionnaire for the self-efficacy of PSTs based on the other dependent variables under consideration could be a starting point. The specificity needs to incorporate characteristics of the chosen pedagogies (or any other chosen content) and language rather than general attitudes towards those with a disability. A starting point for redefining the questionnaire could be the factors noted in the study by Zundans and Lancaster (2012).

Statistical validation of what is described here as a conceptual model using statistical modelling and regression could be investigated to clarify the relationships among the factors associated with inclusion instead of analysing them independently. This study
stepped up from the initial pilot work, now future research could look at large statistical validation of the model based upon larger numbers of students across settings.

A larger scale study could also look at the contribution of other self-organizing principles to the outcomes (Bain, 2007). Schema development could be investigated, for example, where the development of sophisticated planning is enabled to direct PST actions when in classrooms.

Finally, while the predominance of discussion or commentary about the reform of teacher preparation is focused at the policy, standards and structural design level (Arthur-Kelly et al., 2013), there is significant opportunity for more effective approaches to emerge from the bottom up by examining practice in the educational design of courses, and then applying generalisable principles in the classroom context. The program of research, of which this study is a part, seeks to make an emergent contribution to that conversation with a view to improving the quality of teacher preparation overall, starting at the level of course design. Given the high attrition rates experienced by teachers entering the inclusive education field of practice, more effort needs to be expended to assist PSTs to gain the skills and confidence they need to work with the diverse populations of students they will encounter.

Certainly, authors such as Cochran-Smith and Zeichner (2005) have called for more of a shared governance role between universities and school settings. This shared governance could be in the form of the practicum experience or indeed following graduation. These authors noted six major elements that are required in a successful inclusive education PST preparation program: (i) establish a shared language in teacher education curriculum; (ii) support collaboration between special and general educators; (iii) establish collaborative clinical practices; (iv) ensure competence in PSTs before graduation; (v) support beginning teachers in the first three years after graduating; and (vi) finally, promote shared governance that reflects collective responsibility between stakeholders in PST education.

The first three elements from Cochran-Smith and Zeichner (2005) have, in part, been the focus this study and the final two points warrant further investigation. Husebo (2012), for example, suggests the use of action research to support beginning teachers
and to assist with shared governance and collective responsibility. Jimenez-Silva and Olsen (2012) describe a similar construct of community of practice as a ‘teacher-learner community’. After all, “pre-service teachers are preparing for a profession where subject knowledge, and knowledge about a subjects’ didactics, pedagogy and education are inextricably related to teaching practice and the aim of teacher education should be to try to mirror this complex relationship…” Husebo (2012, p. 456).

The following section provides a conclusion to the thesis.

5.10 Conclusions

In conclusion, the results of this study have produced evidence related to four important drivers for inclusive education and teacher preparation.

1. Results demonstrated that pre-service inclusive education teachers can develop the skills required for inclusive practice and also the capacity to communicate that practice in professional ways as they collaborate with others.

2. The present study made an initial contribution to the body of literature that shows the efficacy of teacher education in providing PSTs with those skills.

3. It has established initial support for a design theory for PST courses.

4. Finally, it has made a connection between the skills described in driver 1, the theory and the actual practice of pre-service teachers.

The studies described here stand at the confluence of the four drivers. The results support the contention that the implementation of a theoretically derived course design did covary with the development of pedagogical skills, self-efficacy and pattern language of and those competencies were expressed in actual classroom practice. As such the study contributed to a more robust base of evidence about teacher preparation for inclusion.
References


Costley, K. C. (2013). *Ongoing professional development: The prerequisite for and continuation of successful inclusion meeting the academic needs of special students in public schools*. Online Submission. (ED541075)


Koehnecke, D. (2000). Innovations in teacher preservice programs (pp. 2-8). (ED444957)


Nolte, P. (2010). Using professional development to increase the confidence of teachers to meet the needs of inclusion students. ProQuest LLC. (ED521287)


http://dx.doi.org/10.1080/135697800114177


Appendices
Appendix 1: Course implementation data collection tool across cohort delivery

Strong agreement = (yes I did all of this during the week)
Moderate = (I did SOME of this during the week)
Weak = (I did NOT do this during the week)

<table>
<thead>
<tr>
<th>Week one</th>
<th>Lecture content:</th>
<th>ED cohort</th>
<th>AE cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Strong</td>
<td>Mod</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Strong</td>
<td>Mod</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weak</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Week one PPT … Slides 1-21</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Principles of inclusion Big ideas of inclusion</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Continuum of services and settings that are available</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Rationale for full inclusion verses a continuum option</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Assessment requirements for the subject</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Week one</td>
<td>Tutorial content</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Review of lecture content for inclusion and services available</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Reading objectives</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Data collection:</td>
<td>• SE Questionnaire:</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Knowledge test;</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Demographic information;</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Consent forms etc.</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
### Week two

#### Lecture content:

<table>
<thead>
<tr>
<th></th>
<th>ED cohort</th>
<th>AE cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td>PPT week one slides 22-36</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Two sigma problem and effect size</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Effective characteristics of teaching that can have an impact on student progress in terms of effect size</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>PPT week two slides 1-17</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Collaborative Problem Solving (CPS) - definition</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Steps involved</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Characteristics of CPS</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Need for processes of CPS to be taught</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Need for common language whilst dealing with CPS</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

#### Week two tutorial

<table>
<thead>
<tr>
<th></th>
<th>ED cohort</th>
<th>AE cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td>Review of 2 sigma problem</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Review content from Bloom article</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Review of collaborative problem solving process collaboration workshop**</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Formed own groups.</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Using collaborative problem solving process to solve the problem “name for the group”</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Quiz one reading objectives</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Week three</td>
<td>Lecture content:</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ED cohort</td>
<td>AE cohort</td>
</tr>
<tr>
<td></td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Week three</td>
<td>Review content</td>
<td></td>
</tr>
<tr>
<td></td>
<td>related to ET</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and best practices</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Explicit teaching</td>
<td></td>
</tr>
<tr>
<td></td>
<td>workshop</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discussion about</td>
<td></td>
</tr>
<tr>
<td></td>
<td>quiz expectations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Example lesson</td>
<td></td>
</tr>
<tr>
<td></td>
<td>shown for ET.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week four</th>
<th>Lecture content</th>
<th>ED cohort</th>
<th>AE cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week four</td>
<td>No Lecture this week…. Consolidation</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week four</th>
<th>Tutorial content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ED cohort</td>
</tr>
<tr>
<td></td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Week five Lecture content</td>
<td>Strong</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>• PPT for week five: Cognitive Strategies added to ET.</td>
<td>X</td>
</tr>
<tr>
<td>• Cognitive strategy training / how to:</td>
<td>X</td>
</tr>
<tr>
<td>• Research about cognitive strategies and rationale for use</td>
<td>X</td>
</tr>
<tr>
<td>• Types of CS that may be used.</td>
<td>X</td>
</tr>
<tr>
<td>• Types of learning where CS are suitable to use</td>
<td>X</td>
</tr>
<tr>
<td>• Benefits and limitations of CS training</td>
<td>X</td>
</tr>
<tr>
<td>• Cognitive strategies in the differentiated classroom</td>
<td>X</td>
</tr>
<tr>
<td>Week five tutorial content</td>
<td>Strong</td>
</tr>
<tr>
<td>• Quiz results given back and feedback given. Those who did not reach mastery remained behind to have further discussion and correct their errors.</td>
<td>X</td>
</tr>
<tr>
<td>• Review content related to CS and best practices to use, study guide questions</td>
<td>X</td>
</tr>
<tr>
<td>• Grading rubrics for the CS lesson given out and discussed.</td>
<td>X</td>
</tr>
<tr>
<td>• Example lesson shown for CS.</td>
<td>X</td>
</tr>
<tr>
<td>Week six</td>
<td>ED cohort</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Lecture content</td>
<td>No lecture this week…..consolidation.</td>
</tr>
<tr>
<td>Week six tutorial content</td>
<td>• Review content related to Cognitive Strategies and best practices to use</td>
</tr>
<tr>
<td></td>
<td>• Sharing cognitive strategy lesson plan feedback with peers</td>
</tr>
<tr>
<td></td>
<td>• Discussion about quiz expectations.</td>
</tr>
<tr>
<td></td>
<td>• Grading rubrics for the CS lesson given out and discussed.</td>
</tr>
<tr>
<td></td>
<td>• Example lesson shown for CS.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week seven</th>
<th>ED cohort</th>
<th>AE cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture content</td>
<td>• PPT for week seven: Differentiating ET.</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td>• Differentiation definition</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• The three types of differentiation (process/ product/ content)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• How differentiation works in practice</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• How to modify your learner outcomes?</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• How products can be differentiated?</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• Using cognitive strategies to differentiate</td>
<td>X</td>
</tr>
<tr>
<td>Week seven tutorial content</td>
<td>• Study questions relating to Differentiating ET</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• Differentiating ET workshop and example lesson</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• Quiz 2 preparation and reading objectives from the lecture</td>
<td>X</td>
</tr>
<tr>
<td>Week eight Lecture content</td>
<td>ED cohort</td>
<td>AE cohort</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td>No Lecture … consolidation</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Week eight tutorial content</td>
<td>Reading objectives revisited</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Sharing differentiation Explicit teaching lesson plans for feedback from peers</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Quiz 2 preparation</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Quiz two completed</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week nine Lecture content</th>
<th>ED cohort</th>
<th>AE cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td>PPT for week nine: Cooperative Learning</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Why use Cooperative Learning?</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Theory (cognitive elaboration, motivation theory etc.)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>What is Cooperative Learning?</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>What characteristics make it work?</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Examples of different formats of CL (i.e. Jigsaw 11)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Reading objectives relating to CL</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Grading rubric for cooperative learning (CL)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Cooperative learning workshop including modelling, GP and IP</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Draft CL lesson commenced</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ED cohort</td>
<td>AE cohort</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Week ten</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture content</td>
<td>No Lecture… consolidation</td>
<td></td>
</tr>
<tr>
<td><strong>Week ten tutorial content</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sharing CL lesson plans</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week eleven</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• PPT for week eleven: Differentiating CL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Combining ET and CL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Differentiating of process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Differentiation of product</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Differentiation of content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• CL Peel off</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Implementation cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Week eleven tutorial content</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Reading objectives for differentiating CL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Differentiating workshop with modelling, GP and IP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week twelve Lecture content</td>
<td>ED cohort</td>
<td>AE cohort</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>No lecture… consolidation</td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Week twelve tutorial content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Reading objectives</td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>• Quiz three completed</td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>• Demographic information filled out</td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>• Post questionnaires completed</td>
<td>Strong</td>
<td>Med</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Appendix 2. Component 1 of embedded design: Knowledge and awareness.

Assessment 2: The lesson folio (embedded design ED cohort).

Rationale
The Lesson Folio is the authentic and summative product of student participation in the subject. It is a developmental project informed by ongoing instructor and peer feedback and the students growing knowledge, as reflected in the quizzes and work product completed throughout the subject. The lesson folio builds on the knowledge developed through lecture, reading objectives and quizzes by providing students with the opportunity to apply that knowledge in a series of authentic design tasks.

Task
To submit lessons that employ each of explicit teaching, cooperative learning and cognitive strategy training. Each lesson will be differentiated for different performance levels in the classroom and will include those theoretically derived, research-based characteristics of the practice in question.

Assessment criteria
A grading rubric will be provided to students prior to completion of each of the lessons in the folio. Lessons are submitted on a continuous progress basis on due dates over the course the subject schedule. Each lesson is worth 12% of the subject grade. Students must attain a 70% point performance level on each design to successfully complete this requirement. Where resubmission is required to reach 70% mastery, your first grade will be counted towards your final score for the subject (Subject outline EED212).
The summary paper for the applied experience AE cohort consisted of:
Assessment 2 Summary paper

Rationale
The practical nature of this task allows you to identify, develop, implement and evaluate a program of work linked to the learning needs of the child you have been allocated. The tasks have been designed to provide a link between theory and practice, directly addressing the subject objectives. The Field Site Journal is designed to develop insight, further your understanding and challenge your current conceptions of teaching and learning.

Task
The Summary Paper (2500–3000 words presented in report form) will be informed by the Field Site Journal.

You are to keep a Field Site Journal that documents information about your own learning and teaching as well as the progress and development of your allocated student/s. The Field Site Journal will:

- Provide a general introduction to the child
- Identify the basic learning needs of the child
- Include short weekly lesson plans (day book format) and evaluations
- Include weekly monitoring data
- Describe the reflective and critical analysis of your own learning
- Inform the summary paper about the program and review/examine findings in relation to the research literature.

Assessment criteria

- Your writing needs to reflect your understanding of material covered in lectures and additional readings (minimum of five (5) sources). Your written work will also need to address ethical considerations and specific issues related to cultural sensitivity, privacy and confidentiality of material. Pseudonyms must be used at all times. Present your work in report form, including graphs or other data as relevant (Subject outline EED212).
Appendix 3: Weekly schedule / teaching cycle of the embedded the design principle throughout the content.

<table>
<thead>
<tr>
<th>Week</th>
<th>Lecture</th>
<th>Tutorial</th>
<th>Reading / Notes / Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>Reading objectives</td>
<td>Quiz one reading objectives</td>
</tr>
<tr>
<td></td>
<td>The basics of inclusion and the big ideas</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Collaboration and teamwork</td>
<td>Collaboration workshop</td>
<td>Quiz one reading objectives</td>
</tr>
<tr>
<td>3.</td>
<td>The pedagogy of inclusion explicit teaching</td>
<td>Explicit teaching workshop</td>
<td>Quiz one reading objectives</td>
</tr>
<tr>
<td>4.</td>
<td>Consolidation</td>
<td>Sharing explicit teaching feedback with peers</td>
<td>Explicit Teaching Lesson Draft Due (Five Copies)</td>
</tr>
<tr>
<td>5.</td>
<td>Adapting Explicit teaching</td>
<td>Adaptation Workshop</td>
<td>Final Explicit Teaching lesson due</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Quiz two reading objectives</td>
</tr>
<tr>
<td>6.</td>
<td>Consolidation</td>
<td>Sharing Adapting Instruction Feedback with peers</td>
<td>Draft Adaptation Due (five Copies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Quiz two reading objectives</td>
</tr>
<tr>
<td>7.</td>
<td>Differentiating Explicit Teaching</td>
<td>Differentiation ET Workshop</td>
<td>Final Adaptation Lesson Due</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Quiz two reading objectives</td>
</tr>
<tr>
<td>8.</td>
<td>Consolidation</td>
<td>Sharing Differentiated Explicit Teaching Feedback with peers</td>
<td>Draft Differentiation Due (five Copies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Cooperative Learning</td>
<td>Cooperative Learning Workshop</td>
<td>Final Differentiated Explicit Teaching Lesson due</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Quiz three reading objectives</td>
</tr>
<tr>
<td>10.</td>
<td>Consolidation</td>
<td>Sharing Cooperative Learning Feedback with peers</td>
<td>Draft Cooperative Learning Lesson Due (Five Copies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Quiz three reading objectives</td>
</tr>
<tr>
<td>11.</td>
<td>Differentiating Cooperative learning Lessons</td>
<td>Differentiating workshop</td>
<td>Final Cooperative Learning Lesson Due</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Quiz three reading objectives</td>
</tr>
<tr>
<td>12.</td>
<td>Consolidation</td>
<td>Sharing Differentiated Cooperative Learning Feedback with peers</td>
<td>Draft Differentiated Cooperative Learning Lesson Due (Five Copies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Consolidation</td>
<td>Consolidation</td>
<td>Final Differentiated Cooperative Learning Lesson Due ED cohort &amp; AE cohort</td>
</tr>
</tbody>
</table>

Mid-Session Break
Appendix 4: Pedagogical content knowledge data collection formats

Quiz 1: Pedagogical content knowledge

1. Name at least two principles behind inclusion and explain why each is important in your role as a teacher. (2)
2. What is the 2 sigma problem and what would you do about it? (3)
3. What are the key characteristics of the collaborative problem-solving model? Which are most important for its success and why? (2)
4. What are the steps of the direct teaching model?
5. Why do you think it has achieved such large effect sizes in studies (refer to research in your answer)? (3)
6. How does the correct use of the direct teaching model correctly address a range learners needs? (2)

Quiz 2: Pedagogical content knowledge

1. What does the research say about cognitive strategy training? /1
2. Explain at least one limitation and one benefit of cognitive strategy training. /2
3. How/why would you use cognitive strategies to differentiate instruction? /2
4. What are some examples of ways you can differentiate a learner outcome? Provide an example from your lessons to differentiate: the use of strategy, the content, and the product. /3
5. Describe two types of strategy that could be used to teach “narrative writing”. /2
6. How do you actually teach cognitive strategies to students? /2

Quiz 3: Pedagogical content knowledge

1. Describe the four key characteristics you need to include in CL lessons to make them effective. Why are the details so important? Give an example from research. /2
2. What are improvement points and how do you use them as a reward? How does research support their use in cooperative learning? /1
3. Describe the 3 ways to differentiate instruction. Give an example of each from your own work. /3
4. Pick a theoretical perspective and explain cooperative learning from that perspective. /2
5. Why are the details of CL so important? Give an example from research. /2
6. Explain 2 ways to differentiate a direct teaching lesson? /2
Appendix 5: Decision table for coding of SOLO responses for three quiz occasions

It was expected to see the following for each topic in quiz 1:

- **Inclusion topic**: SOLO /5 possible
- **Collaborative Problem Solving topic**: SOLO /5 possible
- **Explicit teaching topic**: SOLO /5 possible

Decisions to be considered…..

- Terms were only counted once for frequency counts
- SOLO total also calculated as a mean from the three topics covered.

Quiz 2:

- **Cognitive strategies topic**
- **Differentiation**

Decisions to be considered…..

Score of SOLO three must involve multiple entries of points about Cognitive strategies as explicit teaching was assessed in Q1.

Score of /4 involved linking points between CS, or between CS and ET or Differentiation

Quiz 3:

- **Cooperative Learning topic**
- **Differentiation**

Decisions to be considered…..

- Score calculated for Cooperative Learning / 5
- Score also calculated for overall topics that had “inclusion” as an overarching theme or concept. Discussion of overarching theme or concept might refer to “inclusion, students with differing needs, or a variety of students”. The categories that might be included in the discussion included: inclusion, CPS, ET, CL and Differentiation.

Score of /3 could involve multiple points re any of these topics. Core of /4 needed to link them together. A score of /5 needed to link between categories, and connect with theory and/ or research findings.
Appendix 6: Observation measures used for data collection of design and implementation fidelity

Learner Outcome for ET lesson

1. **Condition:** (“Given”….a concept map), Incomplete…..Partially complete….. Complete
   0 1 2
2. **Action:** (What the student will do/learn) Incomplete…..Partially complete….. Complete
   0 1 2
3. **Criterion:** (Specify the level of performance associate with the student learning)
   Incomplete…..Partially complete….. Complete
   0 1 2

Anticipatory Set

4. **Key Words** Absent ……………..Present
   0 1
5. **The attention grabber** Absent ……………..Present
   0 1
6. **Link to Prior Learning(Make a connection)** Absent ……………..Present
   0 1
7. **Model** Absent ……………..Present
   0 1

Guided Practice

8. **GP is broken down into logical steps** Infrequent…Sometimes…. Frequent
   0 1 2
9. **GP relates to the model** Infrequent…Sometimes…. Frequent
   0 1 2
10. **Teacher rotates to check for understanding** Infrequent…Sometimes…. Frequent
    0 1 2

Independent Practice

11. **Students complete all the steps independently** Infrequent…Sometimes…. Frequent
    0 1 2

ET Differentiation
Adapt a learner outcome 3 or 4 ways to include the way you will:

12. **Differentiate the condition,** Incomplete…..Partially complete….. Complete
    0 1 2
13. **Change in the product or action they will complete** Incomplete…..Partially complete….. Complete
    0 1 2
14. **Change the criteria** Incomplete…..Partially complete….. Complete
    0 1 2
Appendix 7: SEIPD data collection formats

Name: ___________________________   Date: _____________________________

Self-Efficacy toward future interactions with students with disabilities SEIPD (Teaching Students).

Below are 15 statements about your potential future professional interactions with students with disabilities. Please circle the response that is most appropriate for you at this time on the scale from Definitely False through Definitely True. These may refer to future teaching of such students.

1. I feel confident in my ability to be able to teach students with disabilities.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

2. I am able to provide individuals /students with appropriate programs.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

3. I can adapt my practices to suit individual needs.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

4. I do not feel in control of any unforeseen situation that may arise during any interaction.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

5. I am confident that I will quickly lose any fear or apprehension.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

6. I do not feel competent in relation to my skills in this area.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

7. When individuals make progress, it is due to the input I have made.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

8. When confronted with a challenging situation I would be likely to give up.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

9. I am able to plan and organise appropriate activities for students with disabilities in my class.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>
10. I am able to attain any goals I set for myself in this area of work.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

11. I have a low expectation of my performance in this area.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

12. I do not look forward to the next time I teach students with disabilities.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

13. It is rare that I feel failure and frustration when working in this area.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

14. These students will benefit greatly from my interactions with them.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>

15. I see my future interactions with students with a disability as successful.

<table>
<thead>
<tr>
<th>Definitely False</th>
<th>False</th>
<th>Mostly False</th>
<th>More False than True</th>
<th>More True than False</th>
<th>Mostly True</th>
<th>True</th>
<th>Definitely True</th>
</tr>
</thead>
</table>
Appendix 8: Demonstration example of an explicit teaching lesson:

Explicit teaching / Direct Teaching Lesson Observation Sheet (example one)

Make note of what you see.....

Learner Outcome for ET lesson
Anticipatory Set
Link to Prior Learning
Model
Guided Practice
Independent Practice

Explicit teaching / Direct Teaching Lesson Observation Sheet (example two)

Make note of what you see.....

Learner Outcome for ET lesson
Anticipatory Set
Link to Prior Learning
Model
Guided Practice
Independent Practice
Appendix 9: SERAP approval

PLANNING AND INNOVATION

Ms Julie Lancaster
2 Kalinda Road
BULLABURRA NSW 2784
AUSTRALIA

Dear Ms Lancaster

SERAP Number 2008/130

I refer to your application to conduct a research project in NSW government schools entitled Theory into practice in Inclusive Education. I am pleased to inform you that your application has been approved. You may now contact the Principals of the nominated schools to seek their participation.

This approval will remain valid until 11 December 2008.

The following researchers or research assistants have fulfilled the Working with Children screening requirements to interact with or observe children for the purposes of this research for the period indicated:

Name Approval expires
Julie Anne Lancaster 16/09/2009

You should include a copy of this letter with the documents you send to schools.

I draw your attention to the following requirements for all researchers in NSW government schools:

- School Principals have the right to withdraw the school from the study at any time. The approval of the Principal for the specific method of gathering information for the school must also be sought.
- The privacy of the school and the students is to be protected.
- The participation of teachers and students must be voluntary and must be at the school’s convenience.
- Any proposal to publish the outcomes of the study should be discussed with the Research Approvals Officer before publication proceeds.

When your study is completed please forward your report marked to General Manager, Planning and Innovation, Department of Education and Training, GPO Box 33, Sydney, NSW 2001.

Yours sincerely

Dr Jenny Donovan
General Manager, Planning and Innovation
24 September 08

NSW Department of Education & Training
Level 6, 35 Bridge Street. GPO Box 33, Sydney, NSW 2001. Tel. 02 8591 6344. Fax 02 8591 8947. E: serap@del.nsw.edu.au
Appendix 10: CSU Ethics Approval

Please note: The protocol number noted here is correct for this thesis. The title differs to the thesis as this is an application to vary the research attached to the same protocol.