THE BIG-FISH-LITTLE-POND EFFECT UNDER THE GRILL:
TESTS OF ITS UNIVERSALITY, A SEARCH FOR
MODERATORS, AND THE ROLE OF SOCIAL COMPARISON

Marjorie Seaton

M.A. (Glasgow University)
B.A. (Psychology Honours) (Macquarie University)

A thesis submitted to the University of Western Sydney in fulfillment of the requirements for the degree of Doctor of Philosophy (Psychology)

October 2007

© M. Seaton 2007
DEDICATION

This thesis is dedicated to my husband and best friend, Willie Seaton, and my children, Shona, Ailsa, and Andrew Seaton.
ACKNOWLEDGEMENTS

I would like to express my sincerest appreciation to Professor Herb Marsh, Professor Rhonda Craven, and Professor Ladd Wheeler for their supervision. All three have given generously of their time, knowledge, and advice. Their teaching has been second to none and their guidance, excellent. Their combined experience has made this journey so much easier for me, and their good humour and patience has made it very enjoyable. Special thanks must go to Professor Herb Marsh. His statistical knowledge is outstanding and I thank him for sharing that knowledge with me. I must also thank him for continuing to supervise me when he moved to Oxford University. Even from 13,000 miles he did not waver in his encouragement, advice, and direction.

I must also thank the researchers who kindly shared their data with me: Hart Blanton, Abraham Buunk, Frederick Gibbons, Hans Kuyper, Pascal Huguet, Florence Dumas, Jean Monteil, Nicolas Genestoux, and Isabelle Regner. To Pascal Huguet and Florence Dumas especially, thank you for your support and assistance. Without the generosity of these individuals, this thesis would not have been possible. My gratitude also goes to all those students who participated in these studies.

Thanks must also go to the administration staff at UWS. To Kate Johnson, Ferina Khayum, and Lisa Car, thank you for all your support. My fellow students deserve special mention. All of you have contributed in some way, from offering statistical advice to helping me grapple with the intricacies of mobile phones and Word!

And lastly, thank you to my family who have been there for me during the highs and lows of my candidature. Without your love, patience, tolerance, and understanding I would never have accomplished this. Thank you all.
STATEMENT OF AUTHENTICATION

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in whole or part, for a degree at this or any other institution.

______________________________
Marjorie Seaton
# TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... X

CHAPTER 1
INTRODUCTION .................................................................................................................. 1

CHAPTER 2
SELF-CONCEPT, THE BIG-FISH-LITTLE-POND EFFECT, AND THE
IDENTIFICATION OF POTENTIAL BFLPE MODERATORS:
BACKGROUND, RESEARCH AND IMPLICATIONS FOR THE
PRESENT INVESTIGATION ................................................................................................. 6

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>6</td>
</tr>
<tr>
<td>A Brief Overview of Historical and Current International Perspectives on</td>
<td></td>
</tr>
<tr>
<td>Academic Selectivity</td>
<td>6</td>
</tr>
<tr>
<td>Historical Background</td>
<td>6</td>
</tr>
<tr>
<td>Examples of Current Policies and Practices</td>
<td>9</td>
</tr>
<tr>
<td>Has Academic Selectivity Produced Expected Benefits?</td>
<td>11</td>
</tr>
<tr>
<td>Section Summary</td>
<td>12</td>
</tr>
<tr>
<td>Theoretical Conceptualisations of the Self-Concept: From Unidimensional to Multidimensional</td>
<td>13</td>
</tr>
<tr>
<td>Historical Conceptualisations</td>
<td>13</td>
</tr>
<tr>
<td>The Shavelson et al. Model</td>
<td>14</td>
</tr>
<tr>
<td>The Marsh and Shavelson Model</td>
<td>15</td>
</tr>
<tr>
<td>Section Summary</td>
<td>19</td>
</tr>
<tr>
<td>The Significance of the Self-Concept Construct</td>
<td>19</td>
</tr>
<tr>
<td>Pervasive Significance of Self-Concept</td>
<td>19</td>
</tr>
<tr>
<td>Research Support for the Importance of a Positive Academic Self-Concept</td>
<td>20</td>
</tr>
<tr>
<td>Section Summary</td>
<td>25</td>
</tr>
<tr>
<td>The Big-Fish-Little-Pond Effect (BFLPE)</td>
<td>25</td>
</tr>
<tr>
<td>Overview of the Big-Fish-Little-Pond Effect</td>
<td>25</td>
</tr>
<tr>
<td>When is the BFLPE Evident?</td>
<td>27</td>
</tr>
<tr>
<td>Early BFLPE Research</td>
<td>28</td>
</tr>
<tr>
<td>Advances in BFLPE Research</td>
<td>29</td>
</tr>
<tr>
<td>Implications of the BFLPE for High-Ability Students</td>
<td>34</td>
</tr>
<tr>
<td>Section Summary</td>
<td>35</td>
</tr>
<tr>
<td>Potential Moderators of the Big-Fish-Little-Pond Effect</td>
<td>36</td>
</tr>
<tr>
<td>Overview</td>
<td>36</td>
</tr>
<tr>
<td>Ability Level as a Moderator of the BFLPE</td>
<td>36</td>
</tr>
<tr>
<td>Socio-Economic Status (SES)</td>
<td>39</td>
</tr>
<tr>
<td>Individual Differences in Learning</td>
<td>40</td>
</tr>
<tr>
<td>Individual Perceptions of the Learning Environment</td>
<td>46</td>
</tr>
<tr>
<td>Section Summary</td>
<td>47</td>
</tr>
<tr>
<td>Implications for the Present Investigation</td>
<td>47</td>
</tr>
<tr>
<td>Generalisability of the BFLPE</td>
<td>47</td>
</tr>
<tr>
<td>Potential Moderators of the BFLPE</td>
<td>49</td>
</tr>
<tr>
<td>Section Summary</td>
<td>53</td>
</tr>
<tr>
<td>Summary</td>
<td>53</td>
</tr>
</tbody>
</table>
CHAPTER 3
RELATING SOCIAL COMPARISON TO THE BIG-FISH-LITTLE-POND EFFECT: A REVIEW OF THE LITERATURE............................................. 55

Introduction ........................................................................................................... 55
Social Comparison Theoretical Perspectives ......................................................... 55
  Pervasive Nature of Social Comparison Processes ........................................... 55
  Festinger’s Social Comparison Theory ............................................................... 56
  Criticisms of Festinger’s Social Comparison Theory ....................................... 57
Section Summary .............................................................................................. 60

Comparison Direction ........................................................................................... 60
  Downward Comparison .................................................................................... 61
  Upward Comparison ......................................................................................... 63
Section Summary .............................................................................................. 69

A Theory of Assimilation and Contrast in Social Comparison – The Selective
Accessibility Model ............................................................................................... 70
  The Selectivity Hypothesis ............................................................................... 71
  The Accessibility Hypothesis ........................................................................... 72
Empirical Research Support for the SA Model....................................................... 72
Revision of the SA Model ..................................................................................... 73
Section Summary .............................................................................................. 74

Generalised Comparison Others ........................................................................... 74
  Historical Conceptualisations of the Generalised Other ................................... 74
Empirical Research Investigating the Generalised Other ........................................ 75
Section Summary .............................................................................................. 77

Comparisons Forced by the Environment.............................................................. 77
  The Distinction Between Selected and Forced Comparisons ........................... 77
Research Evidence for Forced Comparisons ......................................................... 78
Section Summary .............................................................................................. 80

Social Comparison and the BFLPE ...................................................................... 81
  Forced Comparisons with the Generalised Other ........................................... 81
The BFLPE and Low-Ability Students .................................................................. 82
Section Summary .............................................................................................. 83

Implications for the Present Investigation............................................................ 83
  The Blanton et al. (1999) and Huguet et al. (2001) Studies ............................... 83
The Theoretical Basis of the BFLPE – Social Comparison Theory ....................... 84
The Present Investigation ..................................................................................... 84
Summary ............................................................................................................... 87

CHAPTER 4
STUDY 1: TESTS OF THE EXTERNAL VALIDITY OF THE BIG-FISH-
LITTLE-POND EFFECT AND A SEARCH FOR MODERATORS........... 88

Introduction ........................................................................................................... 88
Statement of the Problem ...................................................................................... 89
Background .......................................................................................................... 89
  External Validity of the BFLPE ........................................................................ 89
  A Search for Moderators .................................................................................. 91
Section Summary .............................................................................................. 92
Aims ................................................................................................................ 216
Statement of Hypotheses and Research Questions ....................................... 217
Rationale for Hypotheses and Research Questions ......................................... 221
Methodology ................................................................................................... 223
Results ............................................................................................................. 229
Study 2b Summary .......................................................................................... 247
Discussion ........................................................................................................... 250
Overview ......................................................................................................... 250
Replications..................................................................................................... 250
Comparison Choice......................................................................................... 251
The BFLPE ..................................................................................................... 252
The Selective Accessibility Model.................................................................. 254
Individual Ability ............................................................................................ 255
Strengths and Limitations ............................................................................. 256
Implications for Theory, Research, and Practice ............................................ 258
Summary ............................................................................................................. 260

CHAPTER 6
STUDY 3: SELECTED SOCIAL COMPARISONS AND THE BFLPE .... 262
Introduction ......................................................................................................... 262
Statement of the Problem .................................................................................... 263
Background ......................................................................................................... 263
Aims .................................................................................................................... 264
Statement of the Hypotheses and Research Questions ....................................... 265
Psychometric Properties of Instrumentation ................................................... 265
Tests of the BFLPE ......................................................................................... 267
Comparison Choice......................................................................................... 267
Individual Ability ............................................................................................ 268
Achievement Goal Orientation ........................................................................ 268
Rationale for Hypotheses and Research Questions ............................................. 269
Psychometric Properties of Instrumentation ................................................... 269
Tests of the BFLPE ......................................................................................... 270
Comparison Choice......................................................................................... 271
Individual Ability ............................................................................................ 272
Achievement Goal Orientation ........................................................................ 273
Methodology ....................................................................................................... 275
Participants ...................................................................................................... 275
Measures ......................................................................................................... 276
Design ............................................................................................................. 278
Procedure ......................................................................................................... 280
Results ............................................................................................................. 287
Overview of Analyses ...................................................................................... 287
Psychometric Properties of Instrumentation ................................................... 288
Tests of the BFLPE ......................................................................................... 292
Comparison Choice......................................................................................... 294
Individual Ability ............................................................................................ 299
Achievement Goal Orientation ........................................................................ 302
Discussion ........................................................................................................... 306
Overview ......................................................................................................... 306
Psychometric Properties of Instrumentation ................................................... 306
### LIST OF TABLES

Table 4.1.  *Individualist/Collectivist Index Score by Country (based on Hofstede & Hofstede, 2005)* ........................................................................................................ 110

Table 4.2.  *Economic Classification Based on World Bank (2007) Categories* . 111

Table 4.3.  *Summary of Key Features of the Individual Differences in Learning Indices* ........................................................................................................... 115

Table 4.4.  *Summary of Key Features of Individual Perceptions of the Learning Environment Variables* .......................................................... 118

Table 4.5.  *Summary of Variables Used in Analyses* ........................................... 131

Table 4.6.  *The Effects of Individual Ability and School Average Ability on Math Self-Concept, in a Cross-National Sample* ........................................... 137

Table 4.7.  *The Effects of Individual Ability and School Average Ability on Math Self-Concept, in Individual Countries (Standard Error (SE) in Parentheses)* ........................................... 141

Table 4.8.  *Cultural and Economic Moderators of the BFLPE* ......................... 146

Table 4.9.  *SES Moderators of the BFLPE* ........................................................ 151

Table 4.10.  *Individual Differences in Learning Moderators of the BFLPE* ....... 155

Table 4.11.  *Individual Perceptions of the Learning Environment as Moderators of the BFLPE* ...................................................................................... 166


Table 5.2.  *Number of Classes by Type in Each School* ........................................ 196

Table 5.3.  *Summary of the Constants Added by Class Type* ............................. 197


Table 5.5.  *Recreation of Blanton et al.’s (1999) Data, Regressing T2 Grade on T1 Grade, T2 Self-Evaluation and T2 Comparison Choice, Using Non-Standardised Regression Coefficients* ........................................... 203


Table 5.11.  *Multilevel Modelling Analyses of Blanton et al.’s (1999) Data Predicting Self-Evaluation, with Comparison Choice as a Moderator of the BFLPE* .......................................................... 213


Table 5.15. Individual Grades, Comparison Choice Grades, and Paired t Tests for Huguet et al.’s (2001) Expanded Data .................................................................................................... 228

Table 5.16. Recreation of Huguet et al.’s (2001) Original Data, Regressing T2 Grade on T1 Grade, T2 Self-Evaluation and T2 Comparison Choices 1 and 2, Using Standardised Regression Coefficients ............................................................................ 231

Table 5.17. Recreation of Huguet et al.’s (2001) Original Data, Regressing T2 Self-Evaluation on T2 Grade Comparison Choices 1 and 2, Using Standardised Regression Coefficients ............................................................................ 233

Table 5.18. Multiple Regression Analysis of Huguet et al.’s (2001) Expanded Data, Regressing T2 Grade on T1 Grade, T2 Self-Evaluation and T2 Comparison Choices 1 and 2, Using Standardised Regression Coefficients ............................................................................ 234

Table 5.19. Multiple Regression Analysis of Huguet et al.’s (2001) Expanded Data, Regressing T2 Self-Evaluation on T2 Grade Comparison Choices 1 and 2, Using Standardised Regression Coefficients ............................................................................ 235

Table 5.20. Recreation of Huguet et al.’s (2001) Original Analyses for Grades, using Multilevel Modelling .................................................................................................................... 238


Table 5.22. Multilevel Modelling Analyses of Huguet et al.’s (2001) Expanded Data, Predicting T2 Grades .................................................................................................................... 241


Table 5.24. Multilevel Modelling Analyses of Huguet et al.’s (2001) Expanded Data Predicting Self-Evaluation, with Comparison Choice as a Moderator of the BFLPE .................................................................................................................... 246

Table 5.25. Multilevel Modelling Analyses of Huguet et al.’s (2001) Expanded Data Predicting Self-Evaluation, with Individual Ability as a Moderator of the BFLPE .................................................................................................................... 249

Table 6.1. Summary of Key Features of Scales .................................................................................................................... 279

Table 6.2. Scale Means, Standard Deviations, and Reliability .................................................................................................................... 280

Table 6.3. Summary of Variables Used in Analyses .................................................................................................................... 285

Table 6.4. Individual Ability, Comparison Choice Ability, and Paired t Tests .................................................................................................................... 286

Table 6.5. Individual Ability, Comparison Choice Ability, and Paired t Tests at Levels of Subjective Comparison Choice (Test Scores Standardised) .................................................................................................................... 287

Table 6.6. Results of Confirmatory Factor Analyses for ASDQII and AGQ .... 291

Table 6.7. The BFLPE for French and Math Self-Concept .................................................................................................................... 293

Table 6.8. BFLPE for Non-Choosers for French and Math Self-Concept .................................................................................................................... 293

Table 6.9. Objective Comparison Choice as a Moderator of the BFLPE .................................................................................................................... 295

Table 6.10. Subjective Comparison Choice as a Moderator of the BFLPE .................................................................................................................... 298

Table 6.11. Individual Ability as a Moderator of the BFLPE .................................................................................................................... 300

Table 6.12. Achievement Goals as Moderators of the BFLPE .................................................................................................................... 304

Table C.1. Objective Comparison Choice as a Moderator of the BFLPE when Comparison Choice Test Scores Not Imputed .................................................................................................................... 369
Table C.2. Subjective Comparative Choice as a Moderator of the BFLPE when Comparative Choice Scores Not Imputed (“The Same” is the reference category)........................................................................................................ 370
LIST OF FIGURES

Figure 2.1. The Academic Component of the Shavelson et al. (1976) Model ..... 18
Figure 2.2. The Marsh / Shavelson Model ................................................. 18
Figure 2.3. The Big-Fish-Little-Pond Effect (Marsh & Hau, 2003) ............... 27
Figure 4.1. Predicted Regression Line for a Single-Level Model ............... 123
Figure 4.2. A Random Intercepts Model .................................................... 127
Figure 4.3. A Random Intercepts and Random Slopes Model ................. 128
Figure 4.4. The Relation between Individual Ability and Math Self-Concept,
            Controlling for School-Average Ability .................................... 138
Figure 4.5. Math Self-Concept as a Function of School-Average Ability and
            Individual Ability ........................................................................ 148
Figure 4.6. Higher in Occupation by School-Average Ability Interaction ...... 152
Figure 4.7. Home Education Resources by School-Average Ability Interaction 152
Figure 4.8. Elaboration by School-Average Ability Interaction ................ 156
Figure 4.9. Memorisation by School-Average Ability Interaction ............. 156
Figure 4.10. Extrinsic Motivation by School-Average Ability Interaction ...... 158
Figure 4.11. Intrinsic Motivation by School-Average Ability Interaction ...... 159
Figure 4.12. Math Self-Efficacy by School-Average Ability Interaction ....... 160
Figure 4.13. Math Anxiety by School-Average Ability Interaction ............ 162
Figure 4.14. Preference for Cooperative Learning by School-Average Ability
            Interaction .................................................................................. 164
Figure 4.15. Sense of Belonging by School-Average Ability Interaction ...... 167
Figure 4.16. Attitude to School by School-Average Ability Interaction ....... 167
Figure 5.1. Relation of Individual Ability, Class-Average Ability, and
            Comparison Choice with Self-Evaluation .................................... 189
Figure 5.2. Comparison Other by Class-Average Ability Interaction for Dutch in
            Blanton et al. (1999) ...................................................................... 214
Figure 5.3. Math Ability by Class-Average Ability Interaction in Huguet et al.’s
            (2001) Expanded Data .................................................................. 248
Figure 6.1. Predicted Factor Structure of ASDQ II and AGQ ..................... 266
Figure 6.2. The Effect of the Objective Comparison Choice X Class-Average
            Ability Interaction on French Self-Concept ................................... 296
Figure 6.3. The Effect of the Individual Ability X Class-Average Ability
            Interaction on French Self-Concept ............................................... 301
Figure 6.4. The Effect of the Individual Ability X Class-Average Ability
            Interaction on Math Self-Concept .................................................. 301
Figure 6.5. The Effect of the Performance Avoidance X Class-Average Ability
            Interaction on French Self-Concept ............................................... 305
ABSTRACT

High-ability students will provide future leadership in many fields of endeavour. However, the current practice of educating high-ability students in academically selective classes and schools may not provide the optimum environments for all such students. Big-fish-little-pond effect (BFLPE) research has demonstrated that students in high-ability environments have lower academic self-concepts than equally able students in low-ability settings. Low academic self-concepts are associated with negative educational outcomes. The present investigation aimed to extend current BFLPE theory and research by: (a) testing the BFLPE’s external validity across 41 countries to ascertain the theory’s universality; (b) testing whether the BFLPE is evident in developing countries and in collectivist countries to ascertain whether it transcends cultural and economic barriers; (c) investigating constructs that have the potential to moderate the adverse effects of the BFLPE; (d) elucidating whether upward social comparisons moderate or co-exist with the BFLPE to resolve a conflict in the literature; (e) critically analysing the relation between social comparison processes and the BFLPE to further inform theory; and (f) testing whether the BFLPE varies as a function of ability to clarify its impact on different ability levels. Three studies were conducted to achieve these aims.

Study 1 assessed the BFLPE’s external validity and investigated potential BFLPE moderators. Participants were 265,180 15-year-old students who took part in the 2003 Program for International Student Assessment (PISA) (Organisation for Economic Cooperation and Development, 2005a, 2005b). The BFLPE was evident in the entire cross-national sample of 41 culturally and economically diverse countries, and individually in 38. Moderating effects emerged for socio-economic status (SES), individual differences in learning, and individual perceptions of the learning environment. However, in relation to the large sample these effects were considered small. A larger moderating effect emerged for anxiety, where the BFLPE was found to be greater for highly anxious students.

Study 2 was designed to resolve an apparent contradiction between the social comparison and the BFLPE literatures. Previously published social comparison data (Blanton, Buunk, Gibbons, & Kuyper, 1999; Huguet, Dumas, Monteil, & Genestoux,
2001) were further analysed to ascertain whether the BFLPE was moderated by, or co-existed with, with the beneficial effects of upward comparisons on performance. Participants were 876 Dutch and 1,156 French students. The BFLPE was moderated for Dutch language, but for all other academic subjects it co-existed with selected upward comparisons.

Study 3 assessed the effect of selected comparisons on the BFLPE. Participants were 2,015 French students. The BFLPE was found to co-exist with selected comparisons measured subjectively and when measured objectively for math. The BFLPE was moderated by comparison choice in French when measured objectively and by a performance avoidance goal orientation. The moderating effect of individual ability was tested in all three studies, but provided only small effects, inconsistent in direction.

These results define the BFLPE as externally valid and universally applicable and suggest that it is not moderated by selected social comparisons. Results regarding BFLPE moderators offer practical information to inform intervention strategies that may assist high-ability students to reach their full academic potential.
CHAPTER 1

INTRODUCTION

Throughout the world there is an increasing emphasis on how to educate high-ability students most effectively. High-ability students are expected to be future leaders in fields such as medicine, law, science, and politics. As such, it is critically important that high-ability students are provided with the best education possible to allow them to reach their full potential. The importance of delivering the best education to these students can be argued for on the basis of enriching the intellectual climate of a nation, of strengthening a country’s socio-economic fabric, and of cultivating talent in all facets of society.

Various strategies have been employed to provide quality education for high-ability students including curriculum enrichment, acceleration, and special interest centres (e.g., Frydenberg & O’Mullane, 2000; Merrotsy, 2003). Additionally, students are often selected, on the basis of their academic ability, to attend specially designated classes or schools. However, a growing body of research evidence suggests that segregation on the basis of academic ability does not create optimum environments in which to educate high-ability students (e.g., Craven, Marsh, & Print, 2000; Davis, 1966; Marsh, 1991; 2005; Marsh & Hau, 2003; Marsh & Parker, 1984; Marsh, Trautwein, Lüdtke, & Köller, in review).

At the forefront of this research is the work of Herb Marsh and his colleagues, who have shown that attending academically selective environments has a detrimental effect on students’ perceptions of their academic abilities (a construct referred to in the literature as academic self-concept). This research has consistently found that if students are educated in high-ability classes and schools they will have lower academic self-concepts than their equally able counterparts who are educated in low-ability environments (e.g., Marsh, 1991; 2005; Marsh, Chessor, Craven, & Roche, 1995; Marsh & Hau, 2003; Marsh, Köller, & Baumert, 2001; Marsh et al., in review), a phenomenon known as the “big-fish-little-pond effect” (BFLPE).

The significance of a positive academic self-concept is acknowledged by education policies worldwide as a key outcome of schooling. For example, an aim of
Australian education is that schools should instil in students “qualities of self-confidence, optimism, high self-esteem, and a commitment to personal excellence” (Ministerial Council on Education, Employment, Training, and Youth Affairs, 1999). Moreover, one’s level of academic self-concept can influence important educational outcomes such as course selection, long-term educational aspirations, and academic attainment (e.g., Guay, Larose, & Boivin, 2004; Guay, Marsh, & Boivin, 2003; Hansford & Hattie, 1982; Kelly & Jordan, 1990; Marsh & Köller, 2003; Marsh & Yeung, 1997b). Additionally, academic self-concept has been demonstrated to share a reciprocal causal relation with academic performance (Marsh & Craven, 2006). New reciprocal effects research (Marsh & Craven, 2005, 2006; Valentine & DuBois, 2005) has demonstrated that academic achievement is related to academic self-concept in such a way that they causally impact on each other: High academic achievement is related to improvements in academic self-concept, and high academic self-concept is related to improvements in academic performance such that they are mutually reinforcing. Logically then, for high-ability students to realise their full potential, both academic self-concept and academic achievement must be enhanced simultaneously. If, as BFLPE research shows, attending academically segregated classes and schools lowers academic self-concept, then it follows that high-ability students attending academically selective environments may not be reaching their full academic potential. Given that the effects of academic selectivity on self-concept have been demonstrated to persist for four years after graduation from high school in Germany (Marsh, Trautwein, Lüdtke, Baumert, & Köller, in press), it is of paramount importance to more fully elucidate the BFLPE and its implications for education.

Although BFLPE research has been extensive, there remain areas that are as yet unexplored. Firstly, although there has been considerable cross-national support for the BFLPE (e.g., Craven et al., 2000; Marsh, 1987, 1991, 2004; Marsh et al., 1995; Marsh, Kong, & Hau, 2000; Marsh & Parker, 1984; Mulkey, Catsambis, Steelman, & Crain, 2005; Zeidner & Schleyer, 1998), the countries studied to date have been mostly developed nations and individualist countries. The external validity of the BFLPE in relation to collectivist cultures and less economically developed nations is still to be determined. Hence, the universality of the theory cross-culturally remains to be more fully tested. Secondly, researchers (Marsh, 1991; Marsh et al., 1995) have
proposed that identifying individual differences between students would be a valuable tool in developing policies to minimise the negative effects and maximise the benefits of attending academically selective classes and schools, but to date success in this area has been limited (e.g., Marsh, 1984, 1987, 1991; Marsh & Hau, 2003). As such, little is known about identifying strategies to counteract the effects of the BFLPE for individual students. Finally, researchers who study the BFLPE propose that its theoretical basis lies in social comparison theory (e.g., Marsh et al., 1995; Marsh & Hau, 2003; Marsh et al., 2000). These authors suggest that students compare their achievements with those of their classmates as one means of evaluating their performance. They further speculate that it is these comparisons with classmates, forced on students by the environment in which they find themselves, that lead to lowered academic self-concepts (e.g., Marsh et al., 1995; Marsh & Hau, 2003; Marsh et al., 2000). However, to date this causal link has not been investigated fully in relation to the implicit or explicit comparison processes that individuals actually use. This is unfortunate as such information could assist in informing school-based intervention strategies aimed at countering the negative effects of the BFLPE and allowing students to benefit from academically segregated education.

The purpose of the present investigation was to extend current BFLPE theory and research. More specifically, the broad aims of the present investigation were to: (a) test the external validity of the BFLPE across 41 countries in order to ascertain whether support can be identified for the universality of this theory; (b) test whether the BFLPE is evident in economically developing countries and also in collectivist countries to ascertain whether the BFLPE transcends cultural and economic barriers; (c) investigate potential moderators of the BFLPE to identify constructs that may moderate the adverse effects of the BFLPE; (d) elucidate whether upward social comparisons moderate or co-exist with the BFLPE to resolve a conflict in the literature; (e) critically analyse the relation between social comparison processes and the BFLPE to further inform theory; and (f) test whether the BFLPE varies as a function of ability to clarify the impact of the BFLPE on different levels of student ability. Three synergistic studies that capitalised on advances in statistical methodology were conducted to achieve these aims.
Study 1 was designed to: (a) empirically test the external validity and generalisability of the BFLPE across 41 countries; (b) test if the BFLPE was present across collectivist versus individualist countries and across economically developed versus economically developing countries; and (c) investigate constructs such as individual student characteristics, socio-economic influences, and individual perceptions of the learning environment that could potentially moderate the BFLPE. To achieve these aims, the Program for International Student Assessment (PISA) 2003 database (see Organisation for Economic Cooperation and Development (OECD), 2005a, 2005b), based on data from 41 countries, was utilised to investigate the extent to which the BFLPE varied between countries and whether the size of the BFLPE was moderated by a range of key constructs. Studies 2 and 3 explored the relation between the BFLPE and selected social comparisons, while also examining potential BFLPE moderators. The intent of Study 2 was to determine if selected upward social comparisons, which have been shown to enhance performance, moderate the BFLPE, or co-exist with the BFLPE. This was accomplished through a further analysis of two social comparison studies (Blanton, Buunk, Gibbons, & Kuyper, 1999; Huguet, Dumas, Monteil, & Genestoux, 2001) from a BFLPE perspective, using more powerful statistical techniques. Hence, Study 2 was designed to: (a) replicate the original results of the two social comparison studies; (b) test for the presence of the BFLPE in the same data that demonstrated that selected upward social comparisons enhanced performance; and (c) ascertain whether upward social comparisons moderate or co-exist with the BFLPE, in order to elucidate the relation between the BFLPE and social comparison theory. Study 3 continued the focus on selected social comparisons and their impact on the BFLPE. In particular, Study 3 was designed to: (a) further elucidate the relation between selected social comparisons and the BFLPE by addressing previous limitations through the use of psychometrically sound instrumentation and standardised achievement tests; and (b) test whether individual differences in achievement goal orientation moderate the BFLPE in order to identify characteristics of individuals that may alleviate or aggravate the negative effects of the BFLPE. Additionally, all three studies were designed to investigate the nature of the relation between the BFLPE and individual student ability levels in order to explicate whether the BFLPE varies as a function of individual ability.
In summary, three synergistic studies, using state-of-the-art statistical methodology, were devised to extend BFPLE theory and research and to more fully explicate the relation between BFLPE and social comparison theory. Each study was devised to address limitations of previous research and extend current knowledge of the nature and significance of the BFLPE.
CHAPTER 2


Introduction

The purpose of this chapter is to provide a rationale for the current thesis by reviewing pertinent theoretical and empirical research regarding self-concept, the BFLPE, and potential BFLPE moderators. Firstly, a brief overview of the history of academic selectivity is presented and examples of current policies and practices are outlined. Secondly, some of the more important theoretical models and research in the area are reviewed and the structure and significance of the self-concept construct for optimal academic functioning are discussed. Thirdly, relevant BFLPE research is summarised by highlighting the pervasiveness and generalisability of its effects. Fourthly, potential moderators of the BFLPE are identified on the basis of their importance to other educational domains. Finally, implications for the present investigation arising from research reviewed in this chapter are presented.

A Brief Overview of Historical and Current International Perspectives on Academic Selectivity

Historical Background

The advent of the intelligence testing movement in the early twentieth century brought with it the means of classifying children on the basis of their academic abilities. This made it possible to group students of similar ability together, either at the class or school level, which at the time was espoused as an effective teaching approach to cater for different ability levels (Ireson & Hallam, 2001). As such, in the first half of the twentieth century, many countries around the world espoused ability segregation, in particular selective schooling and ability grouping.
Selective schooling refers to the practice of grouping high-ability students together in specially designated high schools. Examples of these are grammar schools in the UK and selective high schools in Australia. Ability grouping can also occur within schools and can take many forms, but normally it refers to the “separation of same-grade school children into groups or classes that differ markedly in school aptitude” (Kulik & National Research Centre on the Gifted and Talented, 1992, p. 9). In some instances, students are grouped according to ability for specific subjects, such as math (known as setting in the UK or regrouping in the USA); in others, students are grouped homogeneously on the basis of ability for all academic subjects studied (known as streaming in the UK or tracking in the USA; see Ireson & Hallam, 2001).

The popularity of academic selectivity has waxed and waned over the course of the twentieth century in some countries. For example, by the middle of the twentieth century in the UK, selective schools and ability grouping were common practice. Streaming was the typical method of organisation in high schools and in the larger primary schools in the UK (Ireson & Hallam, 2001). After the Second World War, a common test known as the 11+ was introduced for students at the end of their primary education. Performance on this test dictated which secondary school a student would attend. Those who performed best on the test were accepted into grammar schools (also known as senior secondary schools in Scotland); those who performed less well subsequently attended secondary modern schools (also known as junior secondary schools in Scotland). Grammar schools were intended for academically able students who would most likely continue on to higher education; secondary moderns were intended for less able students who would pursue a trade (Emetis, 2007; Secondary Education, 2007).

However, in the UK in the late 1950s the desirability of grammar schools and ability grouping began to be questioned. Intelligent working class children were not gaining access to the selective grammar schools as had been anticipated, the intelligence tests used to select children were queried, and research findings showed negative effects on self-esteem, attitudes, and school engagement (Ireson & Hallam, 2001). Additionally, the political climate in the UK changed and any type of segregation on the basis of academic ability was regarded as discriminatory. The
most equitable system, therefore, was regarded as one that catered for students of all abilities and so comprehensive schools were born. These schools provided both academic and vocational programs for students of all ability levels (Emetis, 2007; Secondary Education, 2007). By the 1980s most selective schools had been abolished and ability grouping had all but been abandoned (Ireson & Hallam, 2001).

Currently 164 grammar schools exist in the UK, but whether or not they will continue to exist is a matter of controversy. Whereas the ruling Labour Party, traditionally, has been opposed to the expansion of grammar schools, the opposition Conservative Party, traditionally, has been their champion. However, recently, the Conservative Party has withdrawn its support for grammar schools, causing a backlash of dissent within the party, their support base, and sections of the community (Qureshi, 2007; Wilson, 2007; Woodhead, 2007).

A similar path was followed in Australia. Following the introduction of intelligence testing in the 1920s, the Directors of Education in each of the six Australian states decided to group high-ability students together in special classes. Thus, in the first half of the twentieth century, a two-tiered system of secondary education was the standard in Australia. Academic pursuits were followed in high schools, while vocational training was provided in technical schools. Academically selective high schools were established in three Australian states for students who were top performers academically (Frydenberg & O’Mullane, 2000). In Australia the late fifties saw the abolition of the two-tier system and the advent of comprehensive education. However, the existing selective schools were allowed to continue within the comprehensive system. As in the UK, the issue of selective schooling in Australia has been a controversial one. Since 1950 reports have recommended that selective schools be phased out (e.g., New South Wales Department of Education, 1977; Vinson, 2002), but as yet none of these recommendations has been carried out, due in large part to the opposition of key stakeholders.

The practice of ability grouping has also been a contentious one in the USA. In that country, since the early 1970s, there has been debate over the educational practice of tracking, whereby students are grouped into three basic categories: (1) academic, (2) average, and (3) slow (Broussard & Joseph, 1998). Some call tracking a form of educational neglect citing its disadvantages (e.g., those in lower tracks are
more likely to drop out of school and engage in antisocial behaviour; see Broussard & Joseph, 1998), while others (e.g., Tieso, 2003) suggest that flexible ability grouping can contribute to “substantial achievement gains both for average and high-ability learners (p. 29). To date, some schools in the USA have detracked, but the debate over the utility of tracking continues.

Examples of Current Policies and Practices

Worldwide, policy documents typically emphasise the importance of allowing every child to reach his/her full potential. However, practice pertaining to academic selectivity and the education of high-ability students differs widely. In the following sections an overview is provided of the approach to the education of high-ability students and academic selectivity currently being employed in a number of countries included in the present investigation, in order to illustrate the diversity of approaches employed internationally.

United Kingdom

In the UK educational policy states that, “every pupil – gifted and talented, struggling or average – should have the right personalised support to reach the limits of their capability” (Department for Education and Skills UK, 2005). The policy continues by stating that high-ability students should be better stretched and challenged within their classes and schools and that they should be given the opportunity to further their talents outside school both locally and nationally.

Today, most schools in the UK are comprehensive, although a small number of state grammar schools and secondary modern schools still exist. Students continue to be placed in these schools on the basis of academic ability (Emetis, 2007; Secondary Education, 2007). Ability grouping continues to be practised. A national survey in the UK (Benn & Chitty, 1996) found that mixed ability grouping had been adopted by about half the schools surveyed for students in their first year of high school. The survey also found that as students progressed through school, ability grouping increased until by Year 9 (third year of high school) most schools practised ability grouping for some academic subjects (see also Ireson, Hallam, Mortimore, Hack, & Clark, 1999).
**Australia**

In Australia, the provision of education is made at a state level. It was not until the 1970s that states produced formal policy documents on the education of high-ability students and even then states approached the issues differently (Robinson, 1992). In South Australia, Tasmania, and Western Australia high-ability students are primarily catered for within each school. For example, the Gifted and Talented Guidelines of the Department of Education and Training in Western Australia state: “Primary and secondary school-based provision occurs within the individual classrooms of all teachers on an ongoing basis. Additionally, schools may develop strategies, which allow the most able students to learn together” (2004, p. 8). Policy in Queensland, New South Wales (NSW), and Victoria allows for selective schools to exist.

In practice today, most students are educated within the comprehensive system, although a small number of selective schools continue to operate. NSW has most selective schools, with 28, and Queensland and Victoria have two each. Ability grouping also occurs at the class level in Australia. Within the comprehensive system, students are segregated according to ability, typically for math and English, and often for additional academic subjects. Moreover, to combat the growing popularity of academically selective schools in NSW, some comprehensive high schools have introduced segregated classes for high-ability students. Formed in the first year of high school, students are segregated into classes on the basis of academic performance in primary school, and continue in these classes for all academic courses. Furthermore, segregation on the basis of ability also occurs at the primary school level in NSW. Not only do “Opportunity Classes” operate in some NSW primary schools for high-ability students in Years 5 and 6, but also students are often segregated into homogeneous classes for math in their last years of primary education.

**France**

French educational policy permits streaming on the basis of academic ability. Students in France attend high school for seven years, but not at the same school. For the first four years they attend a “collège”, which tends to be of mixed ability and in which a core curriculum is taught. For the next three years students attend a “lycee”
or high school, which is streamed according to ability. There are two types of lycée: academic and vocational. Parents, students, and school counsellors together decide which type of lycée a student will attend. Only students who attend the academic lycée can sit the baccalauréat examination that qualifies students to apply for university entrance (Secondary Education, 2007).

French educational policy does not endorse ability grouping at the collège level, although this strategy exists in practice. In France, a collège is a mixed ability lower secondary school, similar to a comprehensive school. According to the “collège unique” act of 1978, no ability grouping is allowed for instructional purposes in these schools. In practice, though, ability grouping appears to be employed in a large number of schools. In a sample of 30,000 students in 212 collèges, Dura-Bellat and Mingat (1998) noted that approximately only 25 percent of students were allocated randomly to classes. These authors concluded that some schools resorted to the practice of ability grouping quite strongly, while others appeared not to use it at all.

The Netherlands

Policy in the Netherlands allows for four levels of secondary education. These levels are graded according to ability. Recommendations are made at the end of primary school, on the basis of school appraisals and standardised tests, as to which level of secondary education a student should attend. The levels range from specialising in training students for the labour market, to preparing students for participation in higher education. Some schools offer a combination of levels; others are more narrowly based and may offer only one level. In combined schools students are often taught in combined classes, regardless of ability, in the first few years of secondary education. However, thereafter, students are grouped depending on the stream to which they have been allocated (Ministry of Education, Culture and Science, 2006).

Has Academic Selectivity Produced Expected Benefits?

It was anticipated that teaching students of similar ability together would help to raise achievement levels. When Ireson and Hallam (2001) reviewed UK and relevant international research evidence in this area, they noted that findings regarding
selective schooling and achievement were mixed, with some studies showing that selective schools produced better examination results. Nevertheless, they cautioned that initial intake ability was often not taken into account, and when it was “the differences between unselective and selective systems are reduced and in some cases disappear” (p. 25). Regarding within school ability grouping, these authors concluded that ability grouping had very little impact on overall achievement.

Ireson and Hallam (2001) also evaluated research examining the impact of ability grouping on self-concept and emotional responses to school. They presented research evidence to show that in schools with higher levels of ability grouping, students had lower self-concepts and less positive attitudes towards the school as a whole and that structured ability grouping had a negative effect on the self-concepts of more able students.

**Section Summary**

In summary, academic selectivity and the education of high-ability students have been controversial issues throughout the twentieth century and into the twenty-first century. Different countries currently approach academic selectivity and the education of high-ability students in diverse ways. Some countries (e.g., the Netherlands and France) have endorsed a more segregated approach to secondary education. Others, such as Australia and the UK, have generally adopted a more comprehensive approach, but nevertheless have kept remnants of the older selective-style schools. However, irrespective of the level of segregation, ability grouping within schools seems to be prevalent. Moreover, academic selectivity does not appear to have delivered the benefits in achievement that were expected of it and indeed has brought with it unexpected adverse side effects in terms of self-concept. It is the issue of self-concept and how it is affected by attending high-ability schools that the remainder of the current chapter addresses.
Having captured the interest of writers for centuries, the notion of the self is one that existed long before the science of psychology was born. Hattie (1992) traced the roots of theorising about the self from Socrates and Plato in the 4th Century BC, through Descartes, James, and Freud, to Bandura and Rogers in the present day: Socrates saw the self as the soul of a person; Descartes separated the mind from the body, with his famous quotation “I think therefore I am”; James contended that the self entails a stream of consciousness; and Rogers argued that “The self is an awareness of being” (Hattie, 1992, p. 34). Notwithstanding the many self theories that exist, there appears to be no formal definition of self-concept among modern-day theorists (Byrne, 1984; Hoge & Renzulli, 1993). However, most would agree that self-concept relates to a person’s sense of self, shaped through interaction with the environment and other people (Shavelson, Hubner, & Stanton, 1976). Thus, how we perceive ourselves is a function of our interactions with others. In a sense others are a mirror through which we catch glimpses of who and what we are (Cooley, 1902). This conceptualisation of self-concept, encompassing both oneself and others, has meant that it has been regarded as a critical variable in a variety of research areas.

For many years there has been vigorous debate among self-concept theorists regarding the structure of the self-concept. Some (e.g., Marx & Winne, 1978; Rosenberg, 1965; Rosenberg & Simmons, 1971) have argued that it is unidimensional in structure, with self-concept (sometimes also referred to as self-esteem, self-appraisal, or self-worth) being denoted by a single construct. For example, Coopersmith (1967) noted that among 10 to 12 year olds, there was no difference in their self-appraisals across different spheres of experience. But, even as far back as 1890, James had regarded self-concept from a multidimensional, hierarchical perspective. He described the self as consisting of three major parts: the material self, the social self, and the spiritual self. He proposed that these three parts of the self are grouped together in a hierarchical structure, with the material self at the bottom, the social self in the middle, and the spiritual self at the top of the hierarchy.
The Shavelson et al. Model

In a seminal work, Shavelson et al. (1976) challenged the view that self-concept had a unidimensional structure. While criticising the area for its lack of sound methodology, measurement instruments, definition, and theory, Shavelson et al. proposed that self-concept could be regarded as both multidimensional and hierarchical in structure. They suggested a possible model in which general self-concept, at the top of the hierarchy, was composed of two second-order factors, academic self-concept and non-academic self-concept (for the academic portion of the model, see Figure 2.1). They theorised that these two factors were further subdivided. According to this model, academic self-concept comprised specific content areas such as English, history, math, and science, whereas non-academic self-concept was divided into social self-concept (relations with peers and significant others), emotional self-concept (particular emotional states), and physical self-concept (physical ability, physical appearance). Within each of these subdivisions, evaluation of behaviour in specific situations was placed at the bottom of the hierarchy. Thus, these authors proposed that self-concept was situation-specific, that one’s self-concept in one domain was dependent on the situation in which one found oneself.

Furthermore, Shavelson et al. (1976) offered a “mature construct definition” of self-concept that was both “formal and explicit” (p. 410). Firstly, they defined the construct of self-concept as a “person’s perception of himself … formed through his experience with his environment … and influenced especially by environmental reinforcements and significant others” (p. 411). They continued by identifying seven characteristics that they proposed were central to this definition. As such, they suggested that self-concept could be depicted as organised, multifaceted, hierarchical, stable, developmental, evaluative, and differentiable. They also emphasised that ideally this definition should be placed within a nomological network, that is, the construct should be shown to have both within-construct and between-construct validity. Within-construct validity relates to the internal structure of the construct, the specific components of which it is comprised, and the structure among these components. An example of within-construct validity is when self-concept is shown to be comprised of different facets, such as academic and non-academic self-concepts. Between-construct validity is demonstrated when the
construct is shown to relate to another independent construct, for example, when academic self-concept is shown to be more strongly related to academic achievement than is physical self-concept (Marsh & Craven, 1997).

Shavelson et al. (1976) did not empirically investigate their model, but rather reviewed the construct validity of five self-concept instruments in use at that time, in regards to their proposed construct definition. Some confirmation was found for their hypothesis that the structure of self-concept was multidimensional and hierarchical, but it could only be regarded as tentative. They found confirmation for a general self-concept from all five instruments and modest support for a multidimensional structure from four of the five. These authors suggested that the subscales could be described as academic, social, emotional, and physical. However, there was little support for the hierarchical structure beyond that of the general self-concept being divided into academic and non-academic self-concept. Whereas the data reviewed suggested that general self-concept was stable, the developmental characteristic of Shavelson et al.’s proposed construct definition could not be tested, and the evaluative feature was not supported. Based on their review, these authors concluded that there was a need for a standardisation of both the definition of self-concept and the instruments used to measure it.

Since then, the multidimensional structure of self-concept has found wide support (e.g., Byrne & Worth Gavin, 1996; Guérin, Marsh, & Famose, 2003; Marsh & Craven 1997, 2006; Marsh, Tracey, & Craven, 2006), whereby numerous factor analyses have demonstrated that self-concept is a multidimensional construct. However, it appears that both the multidimensional and hierarchical structure of self-concept is considerably more complex than initially suggested.

**The Marsh and Shavelson Model**

Marsh and Shavelson (1985) tested the validity of the Shavelson et al. (1976) model by investigating the multidimensional and hierarchical nature of self-concept using the then newly developed Self-Description Questionnaire (SDQ). Developed on the basis of the Shavelson et al. model, this instrument measured seven specific facets of self-concept, such as reading self-concept, physical ability, and peer
relations (for details of this instrument see Marsh, Barnes, Cairns, & Tidman, 1984; Marsh, Relich, & Smith, 1983; Marsh, Smith, & Barnes, 1983). Using confirmatory factor analyses, Marsh and Shavelson found that, for preadolescents, while reading and math self-concepts were each correlated with a general-school facet, the correlation between them was negligible. The model that fit the data best was one that comprised three second-order factors (reading/academic, math/academic, and a non-academic factor) and one third-order factor (general self-concept). As a result, they proposed a revised model, the Marsh/Shavelson model, in which there was no single general academic self-concept; rather two factors, verbal self-concept and math self-concept, were needed to describe relations between the lower-order factors.

Studies of late adolescents reported by Marsh and Shavelson (1985) at the same time, using a version of the SDQ developed specially for that age group (SDQIII), also found that the correlations between the verbal and math components of self-concept were minor, and that they could not be subsumed under a single higher order academic self-concept factor. Whereas there was strong support for the multifaceted structure of self-concept (the SDQIII measures 13 dimensions of self-concept) the hierarchical structure was less clearly delineated for this age group. That the hierarchical structure of self-concept becomes weaker from adolescence onwards has been borne out in subsequent studies (e.g., Byrne & Worth Gavin, 1996; Guerin et al., 2003). Hence, the multidimensional structure of academic self-concept is more complex, and the hierarchical structure much weaker, than originally thought (see Marsh & Craven, 2006, for a review).

Moreover, it appears that academic self-concepts are extremely domain specific. In a subsequent test of the Marsh/Shavelson model, Marsh, Byrne, and Shavelson, (1988) confirmed the existence of two higher order academic factors. However, as the data suggested that two higher-order factors may not be sufficient, these authors proposed that the model was not detailed enough. As a result, further academic subjects, that were judged to be core academic subjects, were added to the model. The new conceptualisation of the academic component of the model is depicted in Figure 2.2. This enhanced model was tested by Marsh (1990b), but in addition to the core subjects represented in that model, other subjects, such as physical education, art, music, and religion, were included. A new instrument was developed to test this
larger variety of academic subjects – the Academic Self-Description Questionnaire (ASDQ) I and II. The ASDQI comprised 13 scales and was tested on students in Grades 5-6; the ASDQII comprised 16 scales and was tested on students in Grades 7-10. Results demonstrated that each scale was able to distinguish the individual factor it was purported to measure, indicating the subject-specificity of academic self-concept. The two-factor higher order structure was adequate to explain the core academic subjects, but for non-core subjects, such as religion, health, art, or music, additional higher order factors were required. Hence, Marsh concluded that when core subjects were considered, these ASDQ studies provided reasonable support for the Marsh/Shavelson model. However, as students were able to distinguish self-concepts in a large variety of different academic subjects, he advised that specific self-concepts should be studied using scales purposely devised to measure that self-concept, and that this caution was even more imperative for researchers studying non-core academic subjects.

Subsequently, the extreme domain specificity of academic self-concept has been validated (e.g., Chapman & Tunmer, 1995, 1997; Lau, Yeung, Jin, & Low, 1999; Marsh, 1992; Marsh, Hey, Roche, & Perry, 1997; Marsh & Roche, 1996; Yeung et al., 2000). For example, Vispoel (1995) administered the SDQIII and the Arts Self-Perception Inventory to university students. Results indicated that self-concepts in dance, dramatic art, visual art, and music skills could be explained by a second-order artistic self-concept factor that was separate from other second-order factors (math, verbal, and non-academic).
Figure 2.1. The Academic Component of the Shavelson et al. (1976) Model

Figure 2.2. The Marsh / Shavelson Model – Based on “A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement” by Marsh, Byrne, and Shavelson, 1988. Journal of Educational Psychology, 80, p. 378.
Section Summary

The self-concept construct is one that has a long and distinguished history. Although historically conceptualised as unidimensional, more recent research has demonstrated that not only is self-concept multidimensional, so too is academic self-concept. As Marsh and Hattie (1996) aptly declared, “there appears to be no support at all for the unidimensional perspective of self-concept or, apparently, even a unidimensional perspective of academic self-concept” (p. 44). In the following sections the importance of a holding a positive self-concept for optimal human functioning is highlighted.

The Significance of the Self-Concept Construct

Pervasive Significance of Self-Concept

Self-concept has been studied across many areas of psychology. It has been examined from a developmental perspective (e.g., Keller, Ford, & Meacham, 1978), cross-culturally (e.g., Markus & Kitayama, 1991), in educational settings (e.g., Craven et al., 2000; Marsh, 1984; 1987; 1990a; 1991; Marsh et al., 1995; Marsh & Hau, 2003; Marsh et al., 2001), in sport and exercise settings (e.g., Fox & Corbin, 1989; Marsh & Fox, 1997), in regards to constructs as diverse as anorexia (Ha, Marsh, & Halse, 2004), bullying (Marsh, Parada, Craven, & Finger, 2004), and winning gold medals in swimming (Marsh & Perry, 2005), and has been perceived as a central ingredient in the development of personality (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2006; Rogers, 1951). It has been studied not only as an important outcome variable in itself, but also as a mediating construct, facilitating other valued outcomes. Humanistic psychologists, like Carl Rogers (1951), argue that a positive self-concept is an extremely important requirement for good mental health. Psychologists, doctors, and teachers all attest to the power of a positive self-concept in overcoming psychological, physical, and learning difficulties. Holding a positive self-concept can aid in improving academic achievement (e.g., Chapman, Tunmer, & Prochnow, 2000; Guay et al., 2003) and prevent one from becoming either a bully or a victim of bullying (Marsh et al., 2004). Additionally, girls with anorexia tend to have lower self-concepts than those without anorexia (Ha et al., 2004). It is little
wonder then that Marsh and Craven (2006) described self-concept as a “hot variable that makes good things happen, facilitating the realization of full human potential in a range of settings” (p. 134).

The development of a positive self-concept is also regarded as a significant and crucial goal of education (Marsh & Hau, 2003; Marsh et al., 2001; Shavelson et al., 1976). Thus, many education systems worldwide include the development and maintenance of a positive self-concept as one of their key objectives. For example, an aim of Australian education is that schools should instil in students “qualities of self-confidence, optimism, high self-esteem, and a commitment to personal excellence” (Ministerial Council on Education, Employment, Training, and Youth Affairs, 1999). Furthermore, one of the main conclusions emanating from a recent OECD study was that student engagement, broadly defined as “students’ attitudes towards schooling and their participation in school activities . . . is closely tied to students’ economic success and long-term health and wellbeing and as such deserves to be treated alongside academic achievement as an important schooling outcome” (Organisation for Economic Cooperation and Development, 2003, p. 9). As such, developing and maintaining a positive self-concept has been acknowledged as an important outcome of education.

**Research Support for the Importance of a Positive Academic Self-Concept**

According to both the Shavelson et al. (1976) and the Marsh/Shavelson (1985) models, academic self-concept, one’s knowledge and perceptions about one’s academic ability (Bong & Skaalvik, 2003), can be considered as one of the principal components of self-concept. As it is such an important element of self-concept, the promotion of a positive academic self-concept can be viewed as an important educational goal in and of itself. Academic self-concept can also be a mediating factor through which other desired outcomes can be achieved. Research has demonstrated that one’s level of academic self-concept can influence factors such as course selection, long-term educational aspirations, educational attainment, academic attainment, and academic achievement (e.g., Guay et al., 2004; Guay et al., 2003; Hansford & Hattie, 1982; Kelly & Jordan, 1990; Marsh & Craven, 2005, 2006; Marsh & Hau, 2003; Marsh & Köller, 2003). The following sections describe
research that illustrates the importance of academic self-concept in facilitating positive educational outcomes and thus allowing students to reach their full potential.

**Course Selection, Aspirations, and University Attendance**

Using a nationally representative longitudinal study of high school students in the United States – the High School and Beyond Study – Marsh (1991) investigated the effect of school-average ability on a wide range of academic outcomes, including coursework selection, educational and occupational aspirations, and university attendance. Three time waves were accessed. The first occurred when students were sophomores (second year of high school), the second, two years later when participants were high school seniors, and the last, two years after graduation from high school. Marsh reported that students in their sophomore and senior years of high school who attended high-ability schools chose less demanding courses, had lower grade point averages, and lower educational and occupational aspirations than students of similar ability who attended low-ability schools. Importantly, the negative effects of school-average ability were still evident for educational and occupational aspirations in the follow-up study, two years after students had left high school. However, academic self-concept had a mediating role. When academic self-concept was controlled, the negative effects of school-average ability were reduced for outcomes in all three time periods (although all but one was still significantly negative). Additionally, academic self-concept had considerable total and direct effects on educational aspirations: The higher a student’s academic self-concept, the more likely it was that the student intended to attend university. Furthermore, although less substantial than the effects on educational aspirations, a higher academic self-concept at Time 2 was associated with university attendance and higher occupational aspirations at Time 3.

Course selection was also investigated in a study by Marsh and Yeung (1997b), who used the ASDQII to measure academic self-concept in nine school subjects. High school students (Grades 8 and 10) were asked to indicate future course selection and to indicate whether they would take the course in the next year and whether they wanted to take it. They found that a high self-concept in a specific subject predicted whether or not students wanted to take that course in the following year and whether they would take it. Interestingly, the relation between school grades
and course selection was not significant in seven of the nine subjects, leading these authors to conclude that self-concepts in specific subjects were more important in the selection of courses than school grades.

**Level of Educational Attainment**

Academic self-concept has also been shown to have an impact on the level of education one achieves. Guay et al. (2004) conducted a 10-year longitudinal study to test the influence of academic self-concept, academic achievement, and family variables, on educational attainment level. Their sample consisted of three cohorts of students from the third, fourth, and fifth grade of primary school. Academic self-concept was measured using the French version of the Self-Perceptions Profile for Children and achievement was measured using a three-item teacher rating scale, which they attested was a valid indicator of achievement. Participants were contacted 10 years later and asked what level of education they had attained. These levels could range from high school, to college, through to university. Correlations between academic self-concept and achievement ranged from .44 for Cohort 1 to .54 for Cohort 3 (both significant), and for academic self-concept and education attainment from .19 to .41 respectively (the correlation for Cohort 1 was not significant). For Cohorts 2 and 3, over and above prior achievement, academic self-concept significantly predicted the educational attainment level of these students 10 years later. For students in Cohort 1 this pattern of results was not significant, but the authors noted that invariance analyses demonstrated that across cohorts the relation between academic self-concept and educational attainment did not differ and was significant, thus lending support for an equivalent relation in all three cohorts. Hence, a positive academic self-concept was associated with better educational outcomes 10 years later. These authors concluded that their findings provided “good support for the long-lasting effects of academic self-concept” (p. 12).

**The Relation Between Academic Self-Concept and Academic Achievement**

Much of the early literature investigating the relation between academic self-concept and academic achievement demonstrated that higher levels of academic self-concept were associated with higher levels of achievement (for a review see Marsh & Craven, 1997). In their meta-analysis, Hansford and Hattie (1982) reported that the average correlation between measures of general self-concept and academic
achievement was moderately low, but positive (.21). However, when measures of academic self-concept were examined, this correlation rose to .42. Similar results were found by Shavelson and Bolus (1982) who reported that the correlation between subject-specific grades and subject-specific self-concepts was stronger than with a more general measure of academic self-concept. Additionally, Marsh (1987) found that, controlling for academic ability and prior achievement, academic self-concept had a causal effect on subsequent academic achievement, albeit a modest one.

Noting the link between academic self-concept and academic achievement, researchers began to question the causal relation between the two: Did a more positive academic self-concept cause higher achievement, did higher achievement cause a more positive academic self-concept, or did each one have a causal effect on the other? This question and its subsequent answer have substantive practical implications for educational interventions and teaching practice. If academic self-concept has a causal effect on academic achievement (known as the self-enhancement model), then interventions, or teacher feedback, based on enhancing self-concept would presumably result in higher achievement. If the opposite were true, and achievement fostered higher academic self-concept (the skill development model), then focusing on ways to improve performance would be the better option (see Calsyn & Kenny, 1977, for details of these models). But, if academic self-concept and academic achievement shared a reciprocal relation, then the superior approach would be to address both academic self-concept and academic achievement simultaneously (Marsh & Craven, 2005, 2006).

Observing that many previous studies examining the causal link between academic self-concept and academic achievement were limited methodologically, Marsh (1990a) conducted what has been referred to as “the classic causal ordering study” (Marsh & Craven, 2005, p. 25). Marsh used data from the Youth in Transition Study, whose participants were boys who attended public high schools in the USA. The study began when these boys were in Grade 10 (Time 1) and finished one year after they had graduated from high school (Time 4). Prior academic ability was measured using scores on four standardised tests at Time 1 only; academic self-concept was measured from responses to three self-rating items at Times 1, 2, and 4;
and academic achievement was measured from reported average grades at Times 1, 2, and 3. Results demonstrated that prior academic self-concept had a significant effect on subsequent academic achievement, after controlling for prior ability. Prior academic achievement (grades) had no effect on subsequent academic self-concept. However, as Marsh and Yeung (1997a) noted, academic achievement at each time point preceded academic self-concept ratings, and as the relations between both were positive, a case could be made for the existence of reciprocal effects.

In later writings, Marsh and colleagues (Marsh, Byrne, & Yeung, 1999; Marsh & Craven, 2005, 2006) argued that the self-enhancement and skills development models were too simplistic and that a more realistic model was one that emphasised that academic self-concept and academic achievement were reciprocally related, whereby high academic achievement is related to improvements in academic self-concept, but in turn high academic self-concept is related to improvements in academic achievement (for a review of these three models see Marsh & Craven, 2005, 2006).

In recent years, support for the Reciprocal Effects Model (REM) has grown steadily (e.g., Guay et al., 2003; Marsh & Köller, 2003; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Muijs, 1997). For instance, Marsh and Yeung (1997a) assessed students’ academic self-concepts and academic achievement across three years. Participants were boys from one Australian high school. The ASDQII was used to measure academic self-concepts for math, English, and science, and academic achievement was measured by standardised school grades. Prior academic achievement had a statistically significant positive effect on subsequent academic self-concepts for all three subjects. Controlling for individual ability, prior self-concepts also had a statistically significant positive effect on subsequent achievement for math, and to a lesser extent for science and English. Hence, this study provided clear support for the REM.

A recent meta-analysis of relevant research further supported the REM, whereby it was concluded that academic self-concept had a significant effect on academic achievement and vice versa (Valentine & DuBois, 2005). This reciprocal relation has also been demonstrated to be valid cross-culturally. For example, Marsh, Hau, and Kong (2002) found support for the REM in a large longitudinal study of high school
students in Hong Kong. Thus, it appears that improvements in academic self-concept lead to higher achievement, and improvements in achievement also lead to higher academic self-concept. These findings have led Marsh and Craven (2006) to advocate that simultaneously improving both academic self-concept and academic achievement may be the optimal method for improving performance.

Section Summary

A positive self-concept is a crucial factor in many areas of human functioning, ranging from maintaining good mental health to winning gold medals. Academic self-concept, a principle component of the self-concept construct, has been shown to be an important factor in producing optimal educational outcomes. A higher academic self-concept can mediate the negative effects of school-average ability and is associated with higher occupational and educational aspirations, university attendance, course selection, and educational attainment levels. In addition to these outcomes, academic self-concept has been shown to be positively related to academic achievement, one of the most important outcomes of education. Furthermore, this relation appears to be reciprocal: A higher academic self-concept is associated with higher academic achievement, and higher achievement is associated with higher academic self-concept. Although academic self-concept plays such a vital role in educational outcomes at the individual level, the promotion of a positive academic self-concept is a goal that has not always been achieved when high-ability students are segregated academically. This failure to fully cater for the needs of high-ability students is discussed further in the following section.

The Big-Fish-Little-Pond Effect (BFLPE)

Overview of the Big-Fish-Little-Pond Effect

The maxim “It is better to be a big frog in a small pond than a small frog in a big pond” was first coined in the psychology realm by Davis (1966), who reviewed the career decisions of college men attending colleges of differing calibre. His findings indicated that high-achieving students had higher ambitions for their careers if they
attended a college where good grades were easy to obtain, than equally good students at a college where the atmosphere was much more competitive. This led him to use the aforementioned maxim to warn parents against sending their sons to the “better” colleges, if there was a chance that the son would be at the bottom of the graduating class.

Since Davis’s (1966) study, much of the research conducted on what has become known as “the big-fish-little-pond effect” (BFLPE), has been undertaken by Marsh and his colleagues (e.g., Craven et al., 2000; Marsh & Hau, 2003; Marsh et al., 2001; Marsh & Parker, 1984; Marsh et al., in press). The BFLPE model that they propose posits that, while individual ability is positively related to academic self-concept, a high school average ability, or a high class-average ability, exerts a negative influence on academic self-concept. Marsh and Hau (2003) define these high-ability schools or classes as places “in which the average achievement level of other students is high” (p. 367). According to this model, one’s academic self-concept partly depends on one’s own ability and partly on the ability of other students in one’s class or school. Moreover, as academic self-concept is domain specific, and as the BFLPE is related to academic self-concept, the BFLPE also appears to be domain specific (see Marsh & Craven, 2006).

An example may help to make this model more concrete. Consider two equally able students who are in their first year of high school. Both were at the top of their year group in primary school, having excelled in school-based tests, standardised state-wide tests, and external academic competitions. Since they both have high-ability, their academic self-concepts are high. However, one student attends the local comprehensive high school, while the other attends an academically selective school in the next suburb. The student in the comprehensive high school is performing well academically and so feels good about his/her abilities. Also, compared to the other students in the school, this student is among the most intelligent, being at the top of the year group (a big fish in a little pond). The student who attends the selective high school is performing around the middle of the year group. There are many other extremely intelligent students at this school, and competition for grades is fierce. Compared to these other students, this student feels that he/she is not very intelligent (a little fish in a big pond). Compared to the student who attends the local
comprehensive school, the academic self-concept of the student who attends the selective school is lower. The selective school has had a negative effect on this student’s academic self-concept (the BFLPE). Hence, while individual ability is positively related to one’s level of academic self-concept, in classes or schools where the ability level is higher, equally able students tend to have lower academic self-concepts than students who are educated in settings where the average ability or achievement levels of classmates is lower. This positive relation between individual ability and academic self-concept and the negative relation between class- and school-average ability and academic self-concept is depicted in Figure 2.3.

![Figure 2.3. The Big-Fish-Little-Pond Effect (Marsh & Hau, 2003)](image)

*Figure 2.3. The Big-Fish-Little-Pond Effect (Marsh & Hau, 2003)*

**When is the BFLPE Evident?**

The BFLPE can exist in both schools and classes where students are segregated by ability. For example, many high schools, whether comprehensive or academically selective, segregate students on the basis of their ability in certain subjects, usually math and English. As discussed above, in recent years comprehensive high schools in NSW, Australia, have introduced an increasing number of high-ability classes for high-ability students in order to combat the growing popularity of academically selective schools. These classes are formed in the early years of high school, with students being grouped according to ability for all academic courses taken in that year. Furthermore, segregation on the basis of ability also occurs at the primary school level in NSW. For example, “Opportunity Classes”, a formal component of
NSW schooling, operate in some NSW primary schools for high-ability students in Years 5 and 6. Hence, a school does not need to be academically selective per se for the BFLPE to exist. Additionally, de facto segregation can occur in relation to socio-economic status (SES). Research suggests that SES is moderately correlated with achievement, whereby the higher the family SES, the higher the individual student achievement (e.g., DeGarmo, Forgatch, & Martinez, 1999; Marks, Cresswell, & Ainley, 2006; Marsh & Parker, 1984; Organisation for Economic Cooperation and Development, 2001a; Pong & Ju, 2000). Thus, wherever students are segregated on the basis of academic ability, and to a lesser extent by SES, the BFLPE may be evident.

**Early BFLPE Research**

*Kulik and Kulik (1982)*

In their meta-analysis investigating the effects of ability grouping on high school students, Kulik and Kulik (1982) located 15 studies that reported self-concept results. Their analyses indicated that the average effect of grouping on self-concept was almost zero, which they considered trivial and concluded that ability grouping had little effect on self-concept. However, Marsh (1984) argued that this conclusion was based on average results and that if different ability groupings (e.g., high-ability students in high-ability schools versus high-ability students in low-ability schools) were considered, then the effect of ability grouping on self-concept would be substantially greater. In support of this argument, Marsh described a study – the Marsh and Parker (1984) study – that was designed to replicate previous research that had displayed a “paradoxically negative correlation between school-average SES and self-concept” (Marsh, 1987, p. 282 – 283). This study was one of the first to provide evidence of the negative effects of school-average achievement on academic self-concept – the BFLPE (but see also Schwarzer, Jerusalem, & Lange, 1983; Strang, Smith, & Rogers, 1978).

*Marsh and Parker (1984)*

These authors assessed the self-concepts of 305 sixth grade Australian students from either high SES schools or low SES schools. Schools were determined to be
either high or low in SES on the basis of local personal knowledge and property values in the area. On average, students in the high SES schools had slightly above average IQ scores ($M = 109, SD = 13.1$), while those in the low SES schools had IQ scores slightly below average ($M = 96, SD = 13.1$). Thus, these students were not explicitly streamed according to ability, but de facto ability grouping occurred as a consequence of differing SES levels. The SDQ was used to measure academic self-concept and individual ability was measured using IQ scores, reading achievement scores, and teacher ratings of academic ability. School-average ability was the average of the individual ability measures for all students within each school. Whereas academic ability had a positive impact on academic self-concept ($\beta = .59$), when individual ability was controlled, the effect of school-average ability on academic self-concept was negative ($\beta = -.27$) – a BFLPE. Equally able students in the high-ability/high SES schools had lower academic self-concepts than students in the low-ability/low SES schools.

**Advances in BFLPE Research**

In the wake of this early research, the BFLPE has found widespread support in a myriad of studies demonstrating the extent of its effects at different levels of education (e.g., Craven et al., 2000; Marsh et al., 1995) and on many educational outcomes (Marsh, 1987, 1991), its durability (Marsh et al., in press) and its generalisability across countries and cultures (e.g., Marsh & Hau, 2003; Marsh et al., 2001; Mulkey et al., 2005; Zeidner & Schleyer, 1998). Examples of such studies are provided in the following sections.

**The BFLPE at Different Levels of Education**

The Marsh and Parker (1984) study demonstrated the existence of the BFLPE at the school level, but the BFLPE has also been shown to exist at the class level. Marsh et al. (1995) conducted a study based on 12 newly established Gifted and Talented classes in primary schools in Sydney, NSW, Australia. Selection for these classes was competitive and was made on the basis of IQ tests and recommendations from teachers and parents. Students from one of these classes (the experimental group) were matched with students from mixed ability classes in neighbouring
schools (the comparison group) on age, gender, and IQ. These students, both male and female, were in Grades 4 to 6, and ages ranged from 9 to 12 years. Students in both the experimental and comparison groups completed a measure of self-concept (the Self Description Questionnaire I) at the beginning of the school year (T1), and then during the second semester (T2). A series of analyses of variance found that students in the experimental group (in the Gifted and Talented class) showed declines in the reading, math, and school components of academic self-concept over time compared to the comparison group (matched students from mixed ability classes). There were no significant differences between the two groups in non-academic self-concept. Thus, this study not only demonstrated that the BFLPE is evident at the class level, but it provided further evidence that it is specific to academic self-concept (Marsh et al., 1995, Study 2). This finding was supported by Craven et al. (2000) who examined the academic self-concepts of students in special Gifted and Talented primary classes, compared to those of gifted students who attended streamed or mixed ability classes. Once again, the students in the special Gifted and Talented classes showed a greater decline over time in their academic self-concepts than the gifted students in the streamed or mixed ability groups. Interestingly though, these authors also assessed student motivation and achievement. Results indicated that for three of six motivational orientations the Gifted and Talented students’ scores were significantly more negative than those of the other two groups, and that there was no difference between the groups on achievement test scores.

However, the BFLPE is not just a symptom of primary school education. The existence of the BFLPE has been repeatedly demonstrated in samples of high school students. For example, studies using seventh graders (Marsh et al., 2001), 10th Grade boys (Marsh, 1987), and students in the senior years of high school (Marsh, 1991), have consistently found a BFLPE.

**Educational Outcomes and the BFLPE**

Apart from negatively affecting academic self-concept, attending high-ability environments has been found to impact on students in other ways. While demonstrating that academic self-concepts were negatively influenced by class-average ability, Marsh (1987), in a re-analysis of the Youth in Transition data,
observed that equally able students in low-ability schools had higher grade point averages than students in high-ability schools. Marsh suggested that this frame of reference effect for grades, although separate, was a contributing factor to the BFLPE for academic self-concept. Additionally, a longitudinal US study of 14,825 high school students provided evidence of a BFLPE across two time periods and demonstrated that school-average ability had a negative effect on a wide-ranging array of educational outcomes (Marsh, 1991). In this study, students in high-ability schools had lower educational and occupational aspirations than students of the same ability who attended low-ability schools. These students were also less likely to take advanced English and math classes, and were more likely to be on a lower academic track than equally able students in lower ability schools.

**Lasting Effects of the BFLPE**

The negative effects of attending a high-ability school also appear to be long lasting. Marsh et al. (in press) conducted two large longitudinal studies examining the long-term stability and persistence of the BFLPE on German high school students. In the first study, 2,306 students were assessed in their final year of high school and then again two years later. Participants at T1 and T2 completed math self-concept items, based on the German version of the SDQIII. Math achievement was measured using items from the Third International Mathematics and Science Study. Results indicated that school-average achievement had a negative effect on math self-concept at T1, the final year of high school, and also at T2, two years after graduation from high school. Furthermore, when the negative effect of school-average achievement on math self-concept at T1 was controlled, school-average achievement still had a statistically significant, but small, negative effect on math self-concept at T2. Study 2 extended these findings by showing that the BFLPE still persisted four years after graduation from high school.

**Generalisability of the BFLPE**

One means of testing the external validity and generalisability of research findings is to establish whether they can be supported in another cultural setting. Matsumoto (2002) suggests that cross-cultural research can inform us whether our psychological constructs can be universally applied. It follows that if psychological constructs can be applied universally across different cultures, then their validity is
greatly strengthened. Although much BFLPE research has been undertaken in Australia (e.g., Craven et al., 2000; Marsh, 2004; Marsh & Parker, 1984; Marsh et al., 1995) and the United States (e.g., Marsh, 1987, 1991; Mulkey et al., 2005), the validity of the findings has also been supported by an impressive validation of the generalisability and universal applicability of the BFLPE by Marsh and Hau (2003), who undertook an extensive study of the BFLPE across 26 countries and by studies in the Middle East, Europe, and Asia. However, when tested experimentally, the generalisability of the BFLPE to different cultural settings has not been supported. In the following section a brief overview of BFLPE research in diverse cultural settings is presented.

Marsh and Hau (2003). Using the OECD’s PISA 2000 database (see Organisation for Economic Cooperation and Development, 2001a, 2001b), Marsh and Hau (2003) investigated the existence of the BFLPE cross-culturally. Their sample consisted of 103,558 15-year-olds from 26 largely OECD countries. Participating countries included the United States, Australia, and countries from Western Europe (e.g., Belgium), Eastern Europe (e.g., Latvia), South America (e.g., Brazil), and Asia (e.g., Korea). Students participating in the PISA study completed a questionnaire that included the three best items from the academic self-concept scale of the SDQII. They also completed achievement tests comprising verbal, math, and science components. Marsh and Hau averaged all available responses for these test items to produce an average achievement score for each individual student. School-average ability was the average of these achievement scores for each school. Using multilevel modelling statistical analyses, results for the entire sample indicated that whereas the effect of individual achievement on academic self-concept was significantly positive, the effect of school-average achievement on academic self-concept was significantly negative. Hence, the BFLPE was evident in the total sample of 26 countries. Marsh and Hau noted that there was variation between the countries and although this variation was small, it was significant. For this reason, they continued by examining the countries individually. As in the entire sample, in all 26 countries the effect of individual achievement on academic self-concept was significantly positive. Conversely, school-average achievement had a negative effect in all 26 countries, and was significantly negative in 24 of the 26 countries. As such,
the results of this study provide strong evidence supporting the cross-cultural generalisability and external validity of the BFLPE.

**Israel.** Zeidner and Schleyer (1998) found support for the BFLPE when they examined academic self-concept, test anxiety, and school grades with 1,020 gifted Israeli primary school children, who were enrolled in either special homogeneous gifted classes, or regular mixed ability classes. Results indicated that gifted Israeli students in academically selective classes had lower academic self-concepts, higher test anxiety, and lower grades than their gifted counterparts in mixed ability regular classes.

**Germany.** In Europe, Marsh et al. (2001) examined math self-concepts among seventh grade East and West German school students as the school systems were being reunited. Prior to unification, the West German students were segregated according to ability, but the East German students were not. After unification both East and West German students were grouped according to ability. This provided researchers with a unique opportunity to examine the development of the BFLPE. These authors found that the negative effect of class-average achievement (the BFLPE) was significantly larger for the West German students at the beginning of the year than for the East German students. However, the difference between the East and West German students had disappeared by the end of the first year of unification, with both displaying a negative BFLPE.

**Hong Kong.** Marsh et al. (2000) examined the BFLPE in Hong Kong, a country known for its collectivist culture, but in which the education system is highly segregated on the basis of ability. Consistent with BFLPE theory and research, they found academic self-concept was lower for students who attended high-ability schools compared to equally able students attending lower ability schools.

**Experimental study by McFarland and Buehler (1995).** An interesting experimental study by McFarland and Buehler (1995, Study 3) provided evidence that the BFLPE may not be as pronounced in collectivist countries. In their study, McFarland and Buehler classified their participants (45 undergraduate students) as having either an individualist cultural heritage or a collectivist cultural heritage on the basis of Hofstede’s (1980) classification system. After completing a bogus test,
half of the participants were told that their performance was high in relation to their group, but that their group as a whole had performed poorly. The other half of the participants were told that they had performed poorly, but their group had performed well. Those who possessed an individualist background reported more favourable reactions to this performance feedback when they thought they were high performers in an unsuccessful group than when they thought they were poor performers in a successful group. The type of feedback had no influence on those from a collectivist heritage. The authors concluded that their findings demonstrated that the BFLPE was reduced for those participants who hailed from collectivist backgrounds as they placed a higher value on their social groups and were less likely to consider their own individual achievements within the group than their individualist counterparts.

**Implications of the BFLPE for High-Ability Students**

Despite the pervasive significance of a positive self-concept for realising one’s full potential, research has shown that students in high-ability schools tend to have lower academic self-concepts when placed in academically selective settings than in heterogeneous settings – the BFLPE (e.g., Craven et al., 2000; Marsh et al., 2001). This effect has important implications specifically for high-ability students, especially when one considers that lower academic self-concepts have been associated with numerous deleterious educational outcomes (e.g., Marsh, 1991). As previously noted, a positive academic self-concept has been linked to coursework selection, long-term educational aspirations, educational attainment, academic attainment, subsequent university attendance, and academic achievement (Guay et al., 2004; Guay et al., 2003; Kelly & Jordan, 1990; Marsh, 1990a; Marsh & Craven, 2005, 2006; Marsh & Hau, 2003; Marsh & Koller, 2003; Marsh & Yeung, 1997a). The validity and generalisability of the consequences of the BFLPE have been demonstrated consistently in numerous empirical studies (e.g., Marsh, 1987; Marsh & Parker, 1984) and cross-culturally in as many as 26 countries (Marsh & Hau, 2003). Additionally, students in high-ability schools tend to select less demanding courses, have lower grade point averages, have lower long-term educational and occupational aspirations, and are on lower academic tracks than equally intelligent students in non-selective schools (Craven et al., 2000; Davis, 1966; Marsh, 1991;
Marsh & Yeung, 1997b). Furthermore, extensive research support for the REM (Marsh & Craven, 2005, 2006; Marsh & Yeung, 1997a; Valentine & DuBois, 2005) demonstrates that academic achievement is related to academic self-concept in such a way that they impact on each other: High academic achievement is related to improvements in academic self-concept, and in turn high academic self-concept is related to improvements in academic performance. Hence, for high-ability students to realise their full potential, academic self-concept and achievement must be enhanced simultaneously. If, as research shows, students in academically segregated schools have lower academic self-concepts, then it follows that they may not be reaching their full potential. Moreover, the effects of the BFLPE have been shown to be long lasting, persisting for at least four years after graduation from high school (Marsh et al., in press). Given that education policies typically emphasise that students should be encouraged to reach the limits of their capabilities (e.g., Department for Education and Skills UK, 2005), that students in high-ability environments may not be doing so, due to lowered academic self-concepts (the BFLPE), is a cause for concern for parents, teachers, and policy makers.

In summary, these correlates of the BFLPE imply that high-ability students in high-ability classes and schools may not be reaching their full academic potential. Hence, it is of paramount importance to evaluate all aspects of the BFLPE to discover how it can be overcome.

Section Summary

Empirical research, dating from 1966, has demonstrated that the effect on academic self-concept of attending a high-ability class or school is negative: Students in high-ability classes and schools tend to have lower academic self-concepts than equally able students in lower ability classes and schools due to the BFLPE. This effect has been shown to exist in classes and schools where students are segregated according to ability, to impact on a number of educational outcomes, to be long lasting, and to generalise across cultures. Moreover, the BFLPE has important implications specifically for high-ability students, not the least of which is that they may not be reaching their full academic potential. The following section addresses
the issue of identifying constructs that may alleviate the adverse effects of the BFLPE for high-ability students.

**Potential Moderators of the Big-Fish-Little-Pond Effect**

*Overview*

Those authors who have demonstrated declines in academic self-concept for students in academically selective schools or gifted and talented programs (the BFLPE) have consistently called for an emphasis to be placed on detecting which students will benefit most, and which will benefit least, from attending these schools and programs (Craven et al., 2000; Marsh, 1991; Marsh et al., 1995). For example, some (e.g., Marsh, 1991; Marsh et al., 1995) have advocated that examining individual differences among students would be a valuable tool in developing policies to maximise the benefits of attending academically selective schools. Although previous research has begun this investigation, success has been limited (Marsh & Hau, 2003; Marsh, 1984, 1987, 1991) and very few constructs have been explicitly examined as BFLPE moderators.

To continue the search for moderators of the BFLPE, the first step is to identify constructs that are important in their own right in other education domains. Constructs that possess a solid theoretical basis are ideal potential candidates to moderate the BFLPE. It is essential that these constructs be identified and examined in this context so that high-ability students are not hampered in their academic pursuits, but are allowed to reach their full potential. In the following sections several potential moderating constructs are identified based on their relevance within other educational domains.

*Ability Level as a Moderator of the BFLPE*

The BFLPE moderator that has received the greatest attention has been individual ability. However, studies investigating whether the BFLPE is similar at all ability levels have provided inconsistent results.
Early Research on Ability Levels and the BFLPE

A study by Coleman and Fults (1985) found that the BFLPE was more pronounced for students of lower ability in academically selective schools and that high-ability students in academically selective classes suffered minimal or no decline in self-concept. However, Reuman (1989) examined math achievement expectancies (as measured by math self-concept, expectancy for success, and perception of math as an easy subject) and math grades for 580 sixth-grade students, who were taught in different types of ability grouping. In some cases, students of the same ability were taught in the same class. Reuman referred to these as homogeneous classes or between-classroom grouping. Others were taught in classes where the ability levels were mixed, but students were grouped according to their ability levels within the classes (heterogeneous classes or within-classroom grouping). Students in both these ability groupings were further subdivided into high, average, or low-achieving groups. Results indicated that compared to the between-class ability grouping, high achievers in the within-class ability grouping had higher achievement expectancies, and low achievers had lower achievement expectancies. To put this in BFLPE terms, students with low-ability suffered more from the BFLPE than high-ability students, but this effect only held for students in the within-classroom ability grouping. Opposite results occurred for those in homogeneous classes. In these classes the high-ability students felt the negative effects of the BFLPE more. Reuman noted that the differences in ability in the homogeneous classes were very small, so presumably students in these classes had very little choice in their comparison targets. These results led Reuman to predict that as classes become more segregated in high schools (i.e., homogeneous between-class groupings become more common), that the math achievement expectancies of high achievers would decline, while those of low achievers would increase.

More Recent Research on Ability Levels and the BFLPE

Marsh et al. (2000) noted that the size of the BFLPE could be affected by individual ability. Although there was some suggestion that higher ability students may suffer less from the BFLPE, the interaction effects of ability and school-average ability were generally small and the directions of the effects varied across models. Similarly, using the nationally representative Youth in Transition Study, Marsh and
Rowe (1996) reported a weak individual ability by school-average ability interaction. An inspection of their data revealed that, although students of all ability levels were affected by the BFLPE, it was students of average ability who suffered the most. However, in a large cross-cultural study of 26 countries, Marsh and Hau (2003) found that, whereas the linear component of individual ability did not moderate the BFLPE, the quadratic component did; although considering the sample size (103, 558), this effect was very small (.008). As a result, these authors concluded that the size of the BFLPE did not vary with individual student ability. Likewise, studies assessing the BFLPE with groups of gifted children in Gifted and Talented classes have demonstrated that the negative effects of attending high-ability classes were similar for all individual ability levels (Marsh et al., 1995).

In summary, in studies where individual ability has been assessed as a BFLPE moderator, results have been contradictory. Some researchers have noted that the BFLPE is more pronounced for students of lower ability in academically selective schools (Coleman & Fults, 1985); others have suggested that between-class ability grouping results in higher academic self-concepts for students of low-ability, and lower academic self-concepts for students of high-ability (Reuman, 1989); while still others have suggested that the BFLPE affects all levels of ability (Marsh et al., 1995; Marsh & Hau, 2003).

**Theoretical Conceptualisations Regarding Individual Ability and the BFLPE**

Not only is there conflicting empirical evidence regarding the importance of individual ability as a moderator of the BFLPE, but different theoretical predictions can also be formulated. On the basis of social comparison theory, one could predict that in high-ability classes and schools, low-ability students would have high-ability students with whom to compare their abilities and these upward comparisons could be a source of inspiration. However, social comparison theory could also predict that these low-ability students would feel inferior when comparing themselves with high-ability same-age peers. Additionally, high-ability students in high-ability classes and schools may have little opportunity to make upward comparisons, as there may be very little difference between the performances of high-ability students (see Reuman, 1989). Moreover, high-ability students in high-ability environments would be able to
compare themselves with low-ability students and presumably these downward comparisons would bolster their self-evaluations.

Conversely, it could be argued that students of all abilities would be affected by the BFLPE. Marsh et al. (1995) have argued that even high-ability students cannot excel in every subject. As students have distinct academic self-concepts for different subjects (see above discussion), students may experience a decline in the academic self-concepts of the subjects in which they do not excel. Moreover, some extremely gifted students may expect to excel in all their subjects and if they do not they may feel very dissatisfied with their overall performance. This being the case, high-ability students may be just as likely to suffer the BFLPE as low-ability students.

**Socio-Economic Status (SES)**

Any research that intends to measure SES is fraught with difficulties, not the least of which is how to operationalise the construct. White (1982) noted that “a wide variety of variables are used as indicators of SES” (p. 462), although income, education, and occupation were, he indicated, the most traditional measures.

**SES and Academic Achievement**

In his meta-analysis investigating the relation between SES and academic achievement, White (1982) concluded that when the individual student was the unit of measurement there was only a weak correlation between SES and academic achievement. Even so, since then studies have demonstrated that higher SES predicts better academic achievement (e.g., Pong & Ju, 2000). In a nationally representative study in 43 countries, home background was found to strongly influence individual student achievement. Parental occupational status, parental education, family wealth, and cultural possessions all had positive associations with academic achievement (Organisation for Economic Cooperation and Development, 2001a). Additionally, research has shown that other factors such as better parenting (DeGarmo et al., 1999) and the possession of cultural resources within the home (Marks et al., 2006) can influence the effect of SES on academic achievement.
**SES and Academic Self-Concept**

Few studies have examined individual SES together with academic self-concept. In their study of sixth grade Australian students described above, Marsh and Parker (1984) selected schools on the basis of socio-economic level, using personal knowledge of the area and local property values. Their sample consisted of 125 students in three high SES schools and 180 students in two low SES schools. These authors demonstrated that individual SES (the average of occupational status and estimated income for each family) had a positive effect on academic self-concept. Bachman and O’Malley (1986) used a sub-sample of the Youth in Transition data to examine the BFLPE. This was a large-scale longitudinal study of young men in the USA conducted from 1966 to 1974. These young men were in Grade 10 when the study began and 23 years of age at its conclusion. Although not directly investigating the effect of SES on academic self-concept, these authors noted that individual SES had a small positive effect on self-concept.

In summary, some studies suggest that higher SES is associated with better academic performance, while others suggest that this association is weak at best. Additionally, the effect of individual SES on academic self-concept has tended to be positive. However, none of the studies reviewed examined whether individual SES had a moderating effect on the BFLPE.

**Individual Differences in Learning**

Students learn differently: Some handle academic tasks with ease, others have difficulty. Individual differences in learning can be seen in the way in which “students address and handle learning tasks in school and the extent to which they are able to achieve their learning goals by applying strategies, motivating themselves, and by controlling and regulating their own learning processes” (Marsh, Hau, Artelt, Baumert, & Peschar, 2006, p. 313).

**Self-Regulated Learning Strategies**

Most teachers and policy makers agree that to learn effectively students need to be self-regulated learners (Boekaerts, 1997). The term “self-regulated learner” is
generally taken to describe students who are “metacognitively, motivationally, and behaviourally active participants in their own learning” (Zimmerman, 1990, p. 4). Zimmerman theorised that students who are self-regulated learners use these metacognitive (e.g., setting goals), motivational (e.g., intrinsic task interest), and behavioural processes (e.g., structuring environments to provide better learning opportunities) to optimise their learning potential. Self-regulated students use strategies that promote effective learning, are motivated to learn, have a belief in their own self-efficacy, and are not troubled by competitive environments (Marsh, Hau, et al., 2006).

Zimmerman (1990) distinguished between self-regulation processes (e.g., intrinsic task interest) and self-regulated learning strategies whose intent is to maximise these processes. Learning strategies include rehearsing and memorising, goal setting and planning, and organising and transforming (for full details of these strategies, see Zimmerman & Martinez-Pons, 1990). Interventions that provide students with the skills to adopt self-regulated learning strategies have proved to be effective (Boekaerts, 1997; Boekaerts & Cascallar, 2006). For example, Camahalan (2006) taught low-achieving mathematics students how to use Zimmerman’s self-regulated learning strategies. Compared to a no-treatment group, these students significantly improved their math performance in subsequent testing.

**Self-Related Cognitions**

**Anxiety.** Many studies have indicated that anxiety and self-efficacy influence academic achievement (Ironsmith, Marva, Harju, & Eppler, 2003). Anxiety has been found to be associated with reductions in grade point averages (Chapell et al., 2005), and to be negatively related to numeracy scores on standardised tests (Martin, 2003). Zeidner and Schleyer (1998) examined the effect of test anxiety on academic performance for gifted students. Two components of test anxiety were assessed: worry and emotionality. Their sample comprised gifted students in two types of educational programs, these being regular mixed-ability classes or special homogeneous classes. Results indicated that test anxiety moderated the effects of the type of educational program on academic performance. Compared to the gifted students in regular classes, test anxiety had a more aversive impact on the academic performance of students in the special homogeneous classes. Meece, Wigfield, and
Eccles (1990) examined the relations between previous math grades, math ability perceptions, performance expectations, and value perceptions on math anxiety. Items used to measure math ability perceptions were similar to those used to assess math self-concept (e.g., “How good are you at math?”). These authors noted that the correlations between their math ability perception items and math anxiety items were consistently negative, ranging from -.11 to -.41.

**Self-efficacy.** Self-efficacy and self-concept can be considered as fundamentally similar, in that it is thought that they both explain and predict thoughts, emotions, and actions. However, depending on the way in which they are operationalised (see Marsh et al., in review), self-concept and self-efficacy can differ in various ways. Self-concept judgements require evaluations of skills and abilities, whereas self-efficacy judgements concentrate on what people believe they can do with the skills and abilities they have. For example, believing that one can successfully punctuate a piece of prose is an efficacy judgement: It is a judgement of how strongly one expects to be able to perform the task successfully. In contrast to academic self-concept then, academic self-efficacy refers to a person’s belief that he/she can succeed in an academic task at specific levels in a particular setting (Bong & Skaalvik, 2003). In addition, while academic self-concept focuses on the past, academic self-efficacy looks forward to the future, and what an individual believes he/she could do. Furthermore, self-concept measures directly imply frame of reference effects, whereas self-efficacy measures do not. In judging one’s academic self-concept the achievements of classmates are used as reference points, but no such reference points are implicated in self-efficacy items.

The relation between academic self-efficacy and academic performance has been well documented. In their meta-analytic review of the literature, Multon, Brown, and Lent (1991) located 38 studies that related academic self-efficacy to academic performance. Their findings indicated a significant positive relation between the two constructs, thus providing evidence that academic self-efficacy facilitates academic performance. Self-efficacy also appears to be positively related to math self-concept. In their assessment of the Student Approaches to Learning Instrument, Marsh, Hau, et al. (2006) noted that self-efficacy and math self-concept were positively correlated (.46).


**Preference for Learning Environment**

The education literature abounds in studies examining cooperative and competitive learning environments. In cooperative classrooms students work together in groups, whereas in competitive environments students’ achievements are evaluated relative to those of other students (Ravenscroft, 1997). Studies, however, tend to produce conflicting results. Some studies conclude that cooperative learning environments produce better academic performance, others that competitive classrooms are superior (Slavin, 1983), while still others have demonstrated that neither is superior. For instance, Sherman (1988) examined achievement in two high school biology classrooms, one using a competitive structure, the other a cooperative one. Both groups completed a test before and after the treatment. Findings indicated that, although both groups scored higher on the post-test than the pre-test, neither the competitive or the cooperative group differed in terms of academic achievement. As regards self-concept, Marsh and Peart (1988) compared a competitive and a cooperative fitness program. Compared to a control group, the competitive program lowered self-concept, while the cooperative program enhanced self-concept. However, when preferences for learning environments have been studied directly, results have indicated that preferences for both types of learning situations (cooperative and competitive) were associated with better performance (Organisation for Economic Cooperation and Development, 2001a).

**Intrinsic / Extrinsic Motivation**

Motivation is considered to be a critical factor in academic achievement (Lepper, 1988; Martin, 2001), and as such, many theories have endeavoured to describe and explain human motivation (for an overview, see Eccles & Wigfield, 2002). Many of these theories characterise motivation as being either intrinsic or extrinsic. Intrinsic motivation has been defined as “the doing of an activity for its inherent satisfaction rather than for some separable consequence” (Ryan & Deci, 2000, p. 56). Individuals who are intrinsically motivated engage in an activity because of the inherent enjoyment they obtain from the activity itself. Students, who are intrinsically motivated to learn, learn because they find the material interesting and enjoyable.

However, as Ryan and Deci (2000) indicate, individuals are not always intrinsically motivated. There are occasions when students have to learn material that
they do not find interesting, but which they must master to obtain a desired outcome (e.g., obtain a school leaving certificate; take a subject that is necessary to finish a desired course). In fact, some would argue that many school subjects fall into this latter category! In this case, students are said to be extrinsically, or instrumentally motivated. Extrinsic motivation is viewed as participating in an activity “in order to obtain some reward or avoid some punishment external to the activity itself” (Lepper, 1988, p. 292). Those who are extrinsically motivated engage in an activity not because they find it pleasurable in and of itself, but because the act of participating provides a desired return.

Whereas intrinsic motivation has been associated with more positive educational outcomes (e.g., Ginsburg & Bronstein, 1993; Gottfried, 1985, 1990), some studies have related extrinsic motivation to poorer outcomes (e.g., Hardre & Reeve, 2003). This was evidenced by Lepper, Corpus, and Iyengar (2005) who examined the association between intrinsic and extrinsic motivation and achievement in third grade through to eighth grade students. They found that students who were intrinsically motivated performed better on standardised achievement tests, but that those who were extrinsically motivated performed worse. Intrinsic motivation has also been associated with higher perceptions of academic competence. Across two studies Gottfried (1985, 1990) offered evidence that students who were more intrinsically motivated perceived themselves to be more academically competent.

However, being extrinsically motivated does not always result in poorer educational outcomes. In their Self-Determination Theory, Ryan and Deci (2000) distinguished between different types of extrinsic motivation: integration, identification, introjection, and external regulation. They proposed that extrinsic motivation could vary in its level of autonomy, with integration being the most autonomous. The more autonomous the motivation: The more free choice one feels one has, the better the outcomes. Supporting evidence for this viewpoint came from a longitudinal study of motivational change in high school students. Otis, Grouzet, and Pelletier (2005) assessed intrinsic motivation, extrinsic motivation, and educational adjustment (homework frequency, dropout intentions, absenteeism, and educational aspirations). Three levels of extrinsic motivation were measured: identification, introjection, and external regulation. Strong positive correlations were found between
intrinsic motivation, identification and educational adjustment. Correlations between introjection, external regulation, and educational adjustment were still positive, but much weaker. However, the correlation between identification and educational adjustment was higher than that for intrinsic motivation, leading these authors to speculate that this type of extrinsic motivation may be especially important within the education environment.

Hence, extrinsic motivation may not be antithetical to intrinsic motivation, and may not always result in poorer educational outcomes. When students have multiple demands to study a lot of material for different exams, the extrinsically motivated students might do better on examinations because they are highly focused on what they need to know to do well on the exam whereas intrinsically motivated students might be more likely to go off on interesting tangents.

Achievement Goal Orientation

The achievement goal approach has also been used to assess achievement motivation. The original model was dichotomous, emphasising mastery and performance goals. Mastery goals have been defined as “students’ desire to learn, improve, and develop competence” (Urdan & Mestas, 2006, p. 354), whereas performance goals have been defined as focussing “on the demonstration of competence relevant to others” (Elliot & Church, 1997, p. 218). Reviews of research on mastery and performance goals in the mid-1990s declared that there was strong evidence for the view that mastery goals resulted in positive outcomes and performance goals resulted in negative outcomes (Elliot, 2005). However, other theorists (e.g., Harackiewicz & Elliot, 1993) questioned the view that performance goals always lead to negative consequences, declaring that there was evidence to show that performance goals could have null or positive effects on some types of achievement (Elliot, 2005).

Recent re-conceptualisations of this model have included a trichotomous achievement goal framework (Elliot & Church, 1997; Elliot & Harackiewicz, 1996) and a 2 X 2 achievement goal framework (Elliot & McGregor, 2001). The former included the distinction between approach and avoidance motivation within the original mastery-performance framework. Whereas the mastery goal remained unchanged in the three-factor version, the performance goal was split into
performance approach (e.g., “It is important for me to do better than other students”, Elliot & McGregor, 2001, p. 504) and performance avoidance (e.g., “I just want to avoid doing poorly in this class”, Elliot & McGregor, 2001, p. 504) goals. A further revision – the 2 X 2 framework – split the mastery goal into mastery approach (e.g., “I want to learn as much as possible in this class”, Elliot & McGregor, 2001, p. 504) and mastery avoidance (e.g., “I worry that I may not learn all that I possibly could in this class”, Elliot & McGregor, 2001, p. 504) goals.

Tests of these models have indicated that whereas mastery goals were positively associated with intrinsic motivation, performance approach goals facilitated achievement. In contrast, performance avoidance goals were negatively related to both intrinsic motivation and achievement (Elliot & Church, 1997; Elliot & McGregor, 2001). For example, Elliot, Shell, Henry, and Maier (2005, Studies 1A and 1B) in tests of the trichotomous model demonstrated that performance in math and verbal intelligence tests was undermined by performance avoidance goals relative to performance approach and mastery goals. These results suggest that performance is facilitated when individuals hold a performance approach orientation.

**Individual Perceptions of the Learning Environment**

The learning environment is “the relatively enduring quality of the school environment that is experienced by participants” (Hoy, Tarter, & Kottkamp, 1991, p. 10). This definition implies that individual perceptions of the learning environment embody constructs such as student-teacher relations, belongingness, and attitudes to school. Various aspects of the learning environment have been associated with a diverse range of outcomes, such as mental health, academic performance, and discipline. In a recent longitudinal study on the impact of violence on adolescents, Ozer (2005) found that school connection, operationalised as “happiness, belonging, safety, closeness, and fair treatment by teachers” (p. 170), was a protective factor in the general mental health of students. Positive student-teacher relations have been shown to facilitate achievement and reduce discipline problems. Crosnoe, Johnson, and Elder (2004) used the National Longitudinal Study of Adolescent Health, a study of American students from Grades 7 to 12, to examine students’ perceptions of their relationships with teachers and how these perceptions related to academic
achievement and discipline issues. Results indicated that, controlling for SES factors and initial individual ability, stronger student-teacher relationships were related to higher academic achievement. Additionally, when SES and prior discipline problems were controlled, stronger student-teacher relationships were associated with fewer subsequent disciplinary problems.

Section Summary

In order to develop policies to maximise the benefits for students of attending academically selective classes and schools, it is necessary to identify constructs that have the potential to alleviate the adverse effects of the BFLPE. Based on their importance in other academic domains, several such constricts were identified, including individual ability, SES, individual differences in learning, and individual perceptions of the learning environment.

Implications for the Present Investigation

Generalisability of the BFLPE

The BFLPE research discussed above suggests that the BFLPE is evident in a number of countries, attesting to the pervasive significance of this effect. However, even though there has been considerable cross-national BFLPE research, it is limited due to its reliance on mainly economically developed countries and individualist nations. Primarily the BFLPE has been studied in Europe, the USA, and Australia. Notable exceptions have been Hong Kong and the cross-national study conducted by Marsh and Hau (2003). Nonetheless, although the Marsh and Hau study was remarkable for the sheer number of countries it examined, there were few economically developing or collectivist countries included. In their sample only five countries could be regarded as collectivist countries (Brazil, Korea, Mexico, Portugal, and Russia; for details of country classification see Hofstede & Hofstede, 2005), and only six as developing (Brazil, Czech Republic, Hungary, Latvia, Mexico, and Russia; see World Bank, 2007). Moreover, Marsh and Hau’s (2003) study combined three subject areas, so was unable to fully explore the domain specificity
of the BFLPE, an important issue considering the multidimensional nature of academic self-concept.

However, it is unclear whether or not developing and collectivist countries will display a BFLPE. For example, compared to developed countries, developing countries tend to have a lower standard of living; they rely more on an agricultural base than an industrial base; poverty is comparably higher; and levels of literacy, education and life expectancy are also lower. In developing countries, where educational standards are lower, there may be no opportunity for the BFLPE to develop, as the academic differences between schools may be minimal. Alternatively, although the education standards on average in these developing countries may be lower than in developed countries, there may be large differences in the education of rich and poor, and as such the education of the rich may well be more reflective of that of developed countries. In this instance, a BFLPE may well be evident in developing countries.

In collectivist countries, the social group to which one belongs is more important than the individual, and the interests, rights and responsibilities of the group take precedence over those of the individual. Children in collectivist societies grow up to revere and respect the members of their extended family and to think of themselves as part of this group. As Hofstede and Hofstede (2005) observed, the group to which one belongs in a collectivist society (usually the extended family) “is the major source of one’s identity” (p. 75). In individualist countries the opposite tends to be true. In these countries the interests, rights, and responsibilities of the individual take precedence over those of the social group to which one belongs. Children in individualist societies tend to grow up in nuclear families, often do not know the members of their extended family, and learn to be independent from a young age. People in individualist societies think of themselves as individuals and their identity is of their own making (Hofstede & Hofstede, 2005). As people in collectivist societies think of themselves in terms of the group and not as individuals, it is possible that the BFLPE will not be evident in such countries. Students in collectivist cultures may place a higher value on the group’s achievements than on their own. If this were the case, then students in high-ability schools in collectivist cultures may
consider the ability level of the school as an asset and assimilate it into their self-concepts, thus eliminating the BFLPE.

Nevertheless, as very few BFLPE studies have been conducted in relation to collectivist or economically developing countries, were support for the BFLPE to be found in these countries, the external validity of the BFLPE would be greatly enhanced. This limited focus of BFLPE research is also problematic given that McFarland and Buehler’s (1995) results suggest that the BFLPE may be reduced for collectivist cultures. However, their research was limited to an experimental study with participants from one country, and the sample size was particularly small. Thus, an important issue for the present investigation to address is whether individuals in collectivist and economically developing countries suffer the negative effects of the BFLPE. To deal with this issue adequately it is necessary to utilise a large sample containing a diverse number of ecologically natural settings. These settings should span both collectivist and individualist cultures and economically developed and developing countries. Additionally, no study to date has explored the domain specificity of the BFLPE in a large cross-national sample, nor whether the BFLPE can be moderated by cultural orientation or a country’s stage of economic development. The present investigation addressed all of these issues in one of the largest cross-cultural BFLPE studies undertaken to date, thereby attending to some of the limitations of previous research.

**Potential Moderators of the BFLPE**

With the exception of individual ability, to date very little research has been conducted in relation to constructs that may alleviate the adverse effects of the BFLPE. The present investigation was designed to overcome this limitation in the literature by identifying potential BFLPE moderators and investigating their effects on the BFLPE. In preceding sections, potential moderators of the BFLPE were identified based on their importance in other educational domains. However, as there has been little research in this area, the ways in which these moderators will affect the BFLPE can be supposition only.
Individual Ability

As regards ability levels, conflicting empirical evidence and competing theoretical perspectives exist regarding individual ability as a moderator of the BFLPE. Thus, a critical concern for the present investigation is to disentangle these contradictory results and viewpoints to clarify the impact of the BFLPE on different levels of student ability.

SES

In relation to SES as a moderator of the BFLPE, research findings are inconsistent. However, they do suggest that SES may be a factor that can influence the BFLPE, although the exact nature of this effect is open to conjecture. Perhaps students from high SES backgrounds may feel more pressure to perform well at school and live up to their parents’ expectations than those from low SES backgrounds and so the BFLPE may be greater for high SES students. Conversely, the BFLPE may be less for high SES students compared to students from low SES backgrounds. High SES students are likely to have parents who are more highly educated. Parents with higher education levels may be more able to provide their children with coping strategies. Hence, high-ability students may be more able to cope with the demands and pressures of high-ability classes and schools than students of low SES backgrounds. As such, compared to students from low SES backgrounds, higher SES students could either suffer more or less from the BFLPE. Hence, whether SES moderates the BFLPE remains to be elucidated by research and as such is an issue addressed in the present investigation.

Individual Differences in Learning

Self-regulated learning strategies. Individual differences in the way students approach learning may also moderate the BFLPE, but as yet these constructs have not been investigated in relation to the BFLPE. For example, students who are self-regulated learners may deal with the school environment in a more positive way than those students who are not. Students who use self-regulated learning strategies may be more confident in their learning as they take a proactive approach to it. The result may be that students who use self-regulated learning strategies may suffer the negative effects of the BFLPE less than those who do not use such strategies.
**Self-related cognitions.** On the basis of previous research in related domains, it seems reasonable to suggest that if students are highly anxious about their academic performance then they may not achieve to their potential. As they are not performing to their best, anxious students may also feel more negative about their abilities, and this may be compounded if they attend a high-ability school. Research examining the relation between math ability perception and math anxiety has also displayed a negative relation and so would support this viewpoint (Meece et al., 1990). Hence, the BFLPE may be more pronounced for students with higher anxiety levels. As regards self-efficacy, if students feel able to succeed academically, and do, then attending a high-ability school or class with high-achieving classmates may have no effect on their self-concepts. Consequently, as self-efficacy also appears to be positively related to self-concept, students high in self-efficacy may not be affected by the BFLPE. Conversely, students in an academically selective class or school who are low in self-efficacy may also have low self-concepts, and because they are surrounded by intelligent students, they may feel that no matter how much effort they put in they will never top the class, thus contributing to the BFLPE. Hence, in high-ability classes or schools, students high in self-efficacy may not suffer the BFLPE to the same extent as those who are low in self-efficacy.

**Preference for learning environment.** Although generally the findings of the effect of type of learning situation on performance are mixed, it could be argued that students who prefer a competitive environment will not be troubled by the competitive nature of high-ability classes or schools: Students who display a high preference for a competitive learning situation may suffer less from the negative effects of the BFLPE. Alternatively, the BFLPE may be greater for these students as they are more likely to compare their achievements with those of fellow students. Students who endorse a preference for learning in a cooperative environment may also suffer the BFLPE to a greater extent, as they may not be able to cope with the competitive nature of high-ability classes and schools.

**Intrinsic / extrinsic motivation.** Intrinsically motivated students may not suffer the negative effects of the BFLPE. If they find academic tasks rewarding in and of themselves, intrinsically motivated students may regard themselves as more capable and may not find the accomplishments of others as threatening, or even relevant, to
their self-views. Implications for extrinsically motivated students are less clear. Some research has demonstrated that extrinsic motivation is associated with negative educational outcomes (Lepper et al., 2005), while other research (Otis et al., 2005; Ryan & Deci, 2000) has shown it to be related with positive educational results. Perhaps if extrinsically motivated students are able to receive the rewards they need to keep them motivated, such as high grades or the praise of teachers and parents, or if they feel a level of autonomy in their behaviour, they also may not suffer the negative effects of the BFLPE. However, in the absence of external rewards, the BFLPE may be greater for extrinsically motivated students.

**Achievement goal orientation.** As mastery goals have been related to intrinsic motivation, this suggests that students in high-ability classes or schools who hold a mastery approach orientation may suffer less from the BFLPE. The focus for mastery approach students is on maximising their learning and so they may not take the performances of others into account when constructing their self-concept. Alternatively, students in high-ability classes or schools who hold mastery avoidance and performance avoidance orientations may suffer more from the BFLPE as their anxiety about their performance may mean that their self-concepts are vulnerable. The effect on the BFLPE of those who hold a performance approach orientation could take either of two directions. As performance approach goals have been associated with higher achievement, these students may take comfort in their accomplishments, with the result that they may not be troubled by the performances of others. Consequently, these students may not suffer the negative consequences of the BFLPE. Alternatively, as the social comparison component of this goal orientation is so high, students may be overwhelmed by comparing themselves with the accomplishments of highly successful others and consequently suffer the negative effects of the BFLPE more strongly.

In summary, individual differences in the way students approach learning, be they self-regulated learners, intrinsically motivated, or highly anxious, may determine their responses to attending high-ability classes and schools. However, no studies of which the author is aware have investigated the relation between individual differences in learning and the BFLPE. Thus, one of the main aims of the
present investigation is to elucidate whether these individual learning differences can moderate the BFLPE.

**Individual Perceptions of the Learning Environment**

Research (e.g., Ozer, 2005) suggests that positive perceptions of the learning environment are related to better outcomes. Perhaps students who perceive their school in a more positive manner may suffer less from the BFLPE. For example, students in high-ability schools may feel more connected to their schools because they do not feel different to other students academically: They are attending a school with many other highly intelligent students and this may promote a sense of belonging. Thus, students who feel more of a sense of belonging to their school may suffer the negative effects of the BFLPE to a lesser extent than those who do not feel so connected. The issue of whether individual perceptions of the learning environment moderate the BFLPE has yet to be elucidated by research and thus is addressed by the present investigation.

**Section Summary**

Two broad implications were drawn from the research reviewed in this chapter. Firstly, to demonstrate the external validity and universal applicability of the BFLPE it is necessary to ascertain its existence in numerous culturally diverse countries. Secondly, to discover ways in which the BFLPE can be alleviated it is necessary to identify and test constructs that have the potential to moderate the BFLPE. The present investigation addressed both of these issues in order to extend present understandings of the BFLPE.

**Summary**

This chapter provided a review of theory and research in relation to self-concept, the BFLPE, and potential BFLPE moderators. It began with a synopsis of the history, current policies, and practices in relation to academic selectivity and the education of high-ability students. Self-concept theory and research were outlined, the significance of self-concept in many facets of life was noted, and the importance of a
positive academic self-concept was emphasised. Relevant BFLPE theory and research were then summarised and implications of the BFLPE for high-ability students in particular were stressed. In addition, constructs were identified that have the potential to moderate the negative effects of the BFLPE. Attention was drawn to implications arising from previous theory and research that are relevant to the present investigation.

The preceding discussion suggests that a positive academic self-concept may not be adequately fostered and developed in an academically selective environment. Since many education systems around the world segregate their students academically, it is important to ascertain the underlying processes that create the BFLPE. Only then can strategies be put in place to alleviate it and increase the academic self-concept of these students. But, what are the processes underlying the BFLPE? Theory underpinning the BFLPE proposes that social comparison processes are responsible for this effect (e.g., Marsh, 1984; Marsh & Hau, 2003). These processes are the focus of the following chapter.
CHAPTER 3

RELATING SOCIAL COMPARISON TO THE BIG-FISH-LITTLE-POND EFFECT: A REVIEW OF THE LITERATURE

Introduction

This chapter relates social comparison theory to the BFLPE. Firstly, social comparison theory is described and critiqued. Secondly, research evidence for upward and downward comparisons is outlined and the selective accessibility theory considered. Thirdly, generalised and forced comparisons are discussed and the relation between social comparison and the BFLPE is explicated. Implications arising from theory and research pertinent to the present investigation are considered. The chapter concludes by providing an overview of the three studies that comprise the present investigation.

Social Comparison Theoretical Perspectives

Pervasive Nature of Social Comparison Processes

Although they may be unaware of the expression, people frequently engage in social comparison. People compare themselves to others in all sorts of ways, from comparing the kind of car they drive with their neighbours’ cars, to comparing their golf handicap with the golf pro’s. People also may make comments based on comparing themselves to others such as: “I look younger than her”; “I’m just as smart as he is”; “I’m a better tennis player.” When such observations are made, people are engaging in social comparison. As social comparison is such a major feature of our everyday lives, there has been an abundance of research endeavouring to understand and explain the psychological mechanisms underlying this element of human behaviour. Underpinning this research is the theory of social comparison processes, developed by Leon Festinger in 1954, which is discussed in the next section.
Festinger’s Social Comparison Theory

In a series of nine hypotheses, eight corollaries, and eight deviations, Festinger (1954) developed a theory to describe and explain how people use others to obtain a sense of their relative standing in order to evaluate their abilities and opinions. For the purpose of the present thesis, only Hypotheses I, II, III, IV, VII, VIII and Corollaries III A and III B are pertinent, and only abilities are discussed here.

The theory’s fundamental proposition is that humans have a drive to evaluate their abilities (Hypothesis I), and when this cannot be done in an objective, non-social way, people compare themselves with others to form these evaluations (Hypothesis II). Festinger (1954) did not detail the motivation behind this drive; he merely mentioned that inaccurate evaluations of ability could “be punishing or even fatal in many situations” (p. 117). Others (e.g., Goethals & Darley, 1977; Gruder, 1977; Jones & Gerard, 1967) have suggested that people are motivated to make comparisons to reduce uncertainty about their abilities. Without being certain about what they can accomplish, people lack the information necessary to make decisions for daily living, and so may make decisions that are not behaviourally adaptive. For example, how does a young runner know whether she is good enough to begin to enter competitions? Will a new resident doctor be able to perform a tricky tracheotomy to save a patient’s life? As Festinger observed, it is often impossible to objectively determine the extent of one’s ability. So, comparisons with others can provide this information and perhaps reduce these uncertainties, thus allowing people to make decisions for everyday living.

Festinger (1954) expanded on these initial hypotheses by describing how people select comparison targets. He postulated that people would cease comparing themselves with others whose abilities were very different from their own (Hypothesis III), especially if they were different on relevant dimensions (Hypothesis VIII). So, according to the theory, our young runner would not compare her running speed with that of a couch potato, or with that of an Olympic champion, because these comparisons would not provide an accurate evaluation of her athletic abilities. A comparison with the former would lead to an overstatement of our runner’s abilities, while a comparison with the latter would result in an understatement. Thus, neither of the comparisons would provide stable and accurate information. Rather,
Festinger theorised that people would prefer to compare themselves with others whose abilities were close to their own (Corollary III A), because this would provide a more accurate evaluation of their abilities than a comparison with someone whose abilities were very different (Corollary III B). Consequently, the theory proposes that comparison targets are more likely to be others similar in ability on the dimensions being evaluated, as they will provide the most accurate evaluations. This “similarity hypothesis” has been the basis of many subsequent studies (e.g., Hakmiller, 1966; Radloff, 1966; Schachter, 1959; Wheeler, 1966), and has been “the most widely cited aspect of the theory” (Suls, 1977, p. 4).

Festinger (1954) also theorised that the usual tendencies to discontinue comparisons with extremely different others might not occur in certain situations. In addition to selecting individuals for comparison purposes, Festinger considered that comparisons could be made with groups (Hypothesis VII) and that situations could arise in which comparisons could be forced on the individual. He suggested that if the group was particularly attractive or if there was no other comparison group available, then the tendency to cease comparing themselves with others whose abilities were very different from one’s own might not occur. Under these circumstances, if the ability could not be changed, Festinger expected, in a foreshadowing of the BFLPE, that the individual would experience “failure and feelings of inadequacy with respect to this ability” (p. 137).

Criticisms of Festinger’s Social Comparison Theory

Related Attributes

Many theorists have noted that Festinger’s Corollary IIIA (people choose someone similar in ability for comparison) and Hypothesis VIII (people cease comparing themselves with someone who is different on attributes related to the dimension being compared) were incompatible (Wheeler, 1991). The former suggests comparisons with others similar on the dimension being assessed; the latter adds to this by suggesting that attributes related to the dimension under evaluation are also considered. As a result, more than 20 years after the theory’s inception, Goethals and Darley (1977) reformulated the similarity hypothesis in an attributional
framework. Their main focus was on the hypotheses concerning the drive to evaluate abilities (Hypothesis I), the preference for selecting similar others as comparison targets (Hypothesis III), and the cessation of comparison with others whose abilities are too divergent from one’s own (Hypothesis VIII). They suggested that people compare their abilities with those of others “who are similar to them on attributes related to performance” (Wheeler & Zuckerman, 1977, p. 336). Consequently, they restated Hypotheses III and VIII as: “Given a range of possible persons for comparison, someone who should be close to one’s own performance or opinion, given his standing on characteristics related to and predictive of performance or opinion, will be chosen for comparison” (p. 265). Thus, when determining how good one’s performance is, the most informative comparison would be with someone who is similar on related dimensions such as age, effort, and practice. Returning to the young runner then, let us suppose that she is 15 years old, can run 400 metres in 64 seconds, and has been training for two years. According to the “related attributes hypothesis” (Wheeler & Zuckerman, 1977, p. 335) a comparison with other 400-metre female runners of a similar age, who have been training for a similar amount of time, would provide her with the most useful information regarding her ability as a runner over this distance. Hence, the related attributes hypothesis suggests that people prefer to compare themselves with others who are similar in related dimensions and whose performance is similar to their own, a hypothesis that has been supported by subsequent research (e.g., Gastorf & Suls, 1978; Wheeler, Koestner, & Driver, 1982; Wheeler, Martin, & Suls, 1997).

**Unidirectional Drive Upward**

Festinger’s (1954) theory has not been devoid of controversy in other respects. In selecting others for comparison, Festinger also theorised that for abilities, “There is a unidirectional drive upward” (p.124). He explained this by stating that, at least in western culture, bettering one’s performance is a highly prized goal. Hence, the theory hypothesises that people will choose comparison targets whose abilities are better than their own. However, the theory does not indicate how the unidirectional drive upward might influence comparison choice (Suls, 1977), nor how it relates to the similarity hypothesis, and these shortcomings have resulted in confusion in the literature as to how the drive upward should be interpreted (Taylor, Buunk, & Aspinwall, 1990; Wood, 1989). Latané (1966) noted that the unidirectional drive
could operate in either an upward or downward direction. Comparing downward may make one feel superior, while comparing upward may provide valuable information. Wheeler (1966) suggested that people compare themselves with slightly superior others, thus combining the unidirectional drive upward with the similarity hypothesis.

These ambiguities in the theory have led to a host of studies designed to determine comparison direction and similarity to the target (e.g., Affleck, Tennen, Urrows, Higgins, & Abeles, 2000; Bellizzi, Blank, & Oakes, 2006; Buunk, Zurriaga, & González, 2006; Gibbons et al. 2002; Thornton & Arrowood, 1966; Wheeler, 1966). Perhaps Festinger did not elaborate on the drive upward because he thought that it might be culturally determined, or because there was little supporting evidence (Wheeler, 1991). Whatever the reason, he has left us with a “masterpiece of ambiguity” (Arrowood, 1986, p. 279), which has spawned much controversy.

**Definition of Social Comparison**

Adding to the ambiguity of the theory is the fact that Festinger did not provide an explicit definition of what constitutes a social comparison. In rectifying this deficit, Wood (1996) suggested that a social comparison consists of multiple processes (acquiring, thinking about, and reacting to social information) with one central component, which she defined as “the process of thinking about information about one or more other people in relation to the self” (pp. 520-521). As regards the processes involved, Wood suggested that the comparer could acquire the social information needed to make a social comparison by actively selecting a comparison target, by constructing or encountering social information, or by a combination of these processes. That social information can be encountered, and that the environment might force comparisons on the individual, was an important departure from the traditional social comparison literature, which had emphasised the active role of the individual in selecting comparisons (Wood, 1989). (This notion of comparisons being forced on the individual will be treated in more detail in a later section). The second process Wood suggested concerns thinking about the social information acquired as regards oneself. Thinking about others, she asserted, does not have to be a conscious act, and involves a search for similarities or differences between the other and oneself on the dimension being evaluated. Reactions to social
comparisons, Wood’s third process, can be cognitive, affective, and behavioural, although effects do not need to be evident for a comparison to have taken place. In recent years, research into the effects of comparisons has mushroomed (e.g., Blanton et al., 1999; Huguet et al., 2001; Mussweiler, 2001a, 2001b, 2003; Stapel & Blanton, 2004; Stapel & Koomen, 2000, 2001, 2005; Stapel & Suls, 2004). In summary, Wood has captured the essence of what it is to make a social comparison in a parsimonious definition and has outlined the component processes involved.

**Section Summary**

Researchers have criticised Festinger’s (1954) theory as being ambiguous regarding its treatment of related attributes and the unidirectional drive upwards, and for the lack of a formal definition of social comparison. While Festinger’s (1954) theory has been criticised in other ways not central to this thesis (see Suls, 1977; Wheeler, 1991), it was remarkable, not only for its logic, insights, and testability, but also because it sparked such a proliferation of diverse research. Topics studied have included: the emotional effects of comparisons (e.g., Buunk, Collins, Taylor, Van Yperen, & Dakof, 1990; Salovey & Rodin, 1984) the different types of comparisons employed (e.g., Buckingham & Alicke, 2002; Suls & Tesch, 1978); the motivational aspects of comparisons (e.g., Lockwood & Kunda, 1997); the behavioural consequences of comparisons (e.g., Stapel & Suls, 2004); and the direction of comparisons (e.g., Collins, 2000; Wills, 1981). Perhaps of all of these, the latter has created the most controversy.

**Comparison Direction**

Doubtless due to the ambiguities surrounding the similarity and unidirectional drive upward hypotheses of Festinger’s (1954) theory, researchers have spent a great deal of time and energy determining the directionality of social comparison processes. However, it was not until 1966, 12 years after the theory’s inception, that any substantial experimental work was undertaken to test the theory’s key assumptions. In that year, the *Journal of Experimental Social Psychology*...
(Supplement 1) published nine experimental studies, most of which examined comparison direction. This supplement has been credited with inspiring much of the subsequent work on social comparison (Suls, 1977), with two papers in particular forever empirically joining “the upward/downward comparison controversy” (Wheeler, 1991, p. 9). These two landmark papers, one by Wheeler demonstrating upward comparison, and the other by Hakmiller demonstrating downward comparison, will be discussed in the relevant sections below.

**Downward Comparison**

*Theoretical Perspectives*

While there was no specific mention of downward comparison in Festinger’s (1954) theory, Latané (1966) speculated that the unidirectional drive could operate in the downward direction. He suggested that an individual might compare him/herself with someone who is worse off, in order to feel better about his/her abilities. Although Festinger considered that the aim of making a social comparison was to provide the comparer with an accurate self-evaluation of his/her abilities, even from the earliest of the social comparison studies there was an indication that accurate self-evaluation may not be the only goal of social comparison (Brickman & Bulman, 1977; Hakmiller, 1966; Thornton & Arrowood, 1966; Wheeler, 1966). Comparing oneself with someone worse off than oneself may not provide an accurate evaluation, but rather an opportunity to self-enhance by protecting and bolstering one’s self-esteem (Wood, 1989).

A landmark paper by Wills (1981) furthered the idea that downward comparisons could be self-enhancing. He proposed that when people experience threat or negative affect, the act of comparing themselves with someone less fortunate could serve to increase their subjective well-being. He continued by asserting that those low in self-esteem would be more likely to engage in downward comparison than those high in self-esteem. In essence, comparing themselves with someone worse off makes people feel better about themselves and their lot in life. In supporting his model Wills reported studies from several areas of social psychology. One of the supporting social comparison studies he cited was the classic study by Hakmiller (1966) in which participants’ egos were manipulated to be under considerable threat. In the
high-threat condition, participants were lead to believe that a personality trait “hostility to one’s parents” was a negative trait. In the low-threat condition, the same personality trait was described more positively. Subsequently, all participants received the same score on the “hostility to one’s parents” personality trait. Compared to those in the low-threat condition, participants in the high-threat condition chose to compare themselves with someone who was even more hostile, and therefore inferior, indicating a downward comparison. Hakmiller concluded that when under threat, choosing to compare oneself with someone less fortunate, rather than a similar other, provided evidence that people use downward comparisons to bolster their self-regard.

Looking specifically at comparisons made under conditions of threat, a number of subsequent studies have produced results consistent with Wills’s downward comparison theory (e.g., Gibbons & Boney McCoy, 1991; Gibbons & Gerrard, 1989, 1991; VanderZee et al., 1996; Wood, Taylor, & Lichtman, 1985). For example, Aspinwall and Taylor (1993) manipulated a threat condition by asking participants to recall a recent academic setback. Participants completed self-esteem and mood measures, and then read either about a high (upward comparison) or a low (downward comparison) achieving student. Scales assessing post-comparison affect and adjustment to college were subsequently completed. Those in the control group did not read about a student, but completed all other measures. They found that participants who were low in self-esteem evaluated themselves more favourably after being exposed to the downward comparison information than those in the control group. These authors concluded that, “downward social comparisons can bolster self-evaluation … in the comparison domain under conditions of naturalistic threat” (p. 718). This finding, that it is those low in self-esteem who benefit most from downward comparison, is consistent with Wills’s (1981) theory. Moreover, in a large-scale longitudinal study of people with cancer, VanderZee et al. (1996), using a control group of healthy people, found that cancer patients had a higher need for social comparison, compared with less fortunate others more often than the healthy controls did, and felt better off than others after comparing downwards.
**Criticisms of Downward Comparison Theory**

Researchers have, nevertheless, questioned the basic tenets of downward comparison theory, namely, that one compares downward when unhappy, threatened, or low in self-esteem (e.g., Buunk, Van Yperen, Taylor, & Collins, 1991; Taylor & Lobel 1989; Ybema & Buunk, 1993). In a naturalistic study of everyday comparisons, Wheeler and Miyake (1992) asked students, over a two-week period, to keep a diary of the comparisons they made, the direction and dimension of the comparison, and their mood. Two of their findings in particular were contrary to predictions from downward comparison theory (Wills, 1981): (1) When their participants experienced negative affect they compared upward, but when affect was positive they compared downward; and (2) on the Lifestyle dimension, it was those who were higher in self-esteem who made more downward comparisons. These results led Wheeler and Miyake to suggest that perhaps the basic principles of downward comparison theory were erroneous. They claimed that the research most frequently cited in the theory’s favour did not provide evidence convincing enough to show that threat or negative affect results in downward comparison. Subsequent research has confirmed these conclusions, by demonstrating that downward comparisons occur when participants are happy (Affleck et al., 2000; Wood, Michela, & Giordano, 2000), and that individuals make upward comparisons even under threat (Lockwood, Dolderman, Sadler, & Gerchak, 2004; VanderZee, Oldersma, Buunk, & Bos, 1998; Ybema & Buunk, 1993).

**Upward Comparison**

**Upward Comparison with Similar Others**

While there has been very little support for downward comparison as a way of accurately evaluating one’s abilities, research has consistently indicated that when people wish to evaluate their abilities, they do so by comparing themselves with similar others who are performing slightly better than themselves (e.g., Arrowood & Friend, 1969; Collins, 1996, 2000; Friend & Gilbert, 1973; Gruder, 1971; Thornton & Arrowood 1966; Wheeler et al., 1969; Wheeler & Koestner, 1984; Wood, 1989).
In one of the early social comparison studies, Wheeler (1966) introduced the rank order paradigm to examine comparison direction under differing motivational conditions. Participants performed a task on which they were given false feedback and a rank ordering of the other participants’ performances. They were told that they could look at another participant’s score associated with one other rank. Most participants chose to compare themselves with someone whose score was one rank above their own, even in the low motivation condition. Thus, participants preferred to compare themselves with someone whose rank was similar to their own, but slightly better.

In a naturalistic setting, Nosanchuk and Erickson (1985) used hypothetical scenarios to assess the direction of comparisons made by bridge players. The participants were asked to nominate whether they would prefer to compare their bridge playing ability with someone similar, better, or inferior in ability. Participants indicated that they preferred to compare with similar others. But, Nosanchuk and Erickson also asked their participants to actually name their comparison targets. In doing so, these authors were able to gauge objectively the playing ability of the nominated target. Using this objective measure of comparison level, they found that their participants preferred to compare their bridge-playing ability with someone who was in fact a better player. Hence, participants had perceived the superior players as similar in ability to themselves.

**Upward Comparison and Contrast Effects**

Although early researchers (Thornton & Arrowood, 1966; Wheeler, 1966) suggested otherwise, historically the social comparison literature has regarded the consequences of social comparisons, including upward comparisons, as being typically contrastive in nature: Individuals evaluate themselves more negatively after comparing themselves with superior others and more positively after comparing themselves with inferior others (e.g., Brickman & Bulman, 1977; Collins, 1996, 2000; Major, Testa, & Bylsma, 1991; Suls & Wheeler, 2000; Wills, 1981).

Probably the best and most oft-cited example of the aversive effects of upward comparisons is a study by Morse and Gergen (1970). While they were waiting to be interviewed, ostensibly for a summer job, Morse and Gergen’s participants were exposed to either a very well groomed fellow applicant (Mr. Clean) or a poorly...
dressed fellow applicant (Mr. Dirty). Self-esteem was measured before and after the encounter. Those who were confronted by Mr. Clean (and who presumably compared upward) showed a decline in self-esteem, while the self-esteem of those who were exposed to Mr. Dirty (and who compared downward) increased. Such results lend credence to the notion that upward comparisons can be painful.

Other studies have also demonstrated the negative effects of upward comparisons. In lieu of a no-comparison control group, Cash, Cash, and Butters (1983) used an extremely dissimilar group. In addition to presenting their all-female participants with photographs of attractive and unattractive women, they also showed them photographs of professional models. The professional model photographs were the same as those used for the attractive condition except that by attaching the advertiser’s name to the photographs it was made explicit that they were professionals. The authors argued that professional models would be less appropriate comparison targets and so participants would judge themselves as less attractive after viewing photographs of attractive women as opposed to professional models. Participants were randomly assigned to one of three conditions depending on which photographs they viewed: attractive, not attractive, and professional models. When participants viewed pictures of attractive women they rated their own attractiveness lower than those who looked at pictures of non-attractive women. Additionally, those in the professional model condition rated themselves as more attractive than those in the attractive condition. Although a no-comparison condition would have been more useful in clearly delineating the direction of the comparisons, the professional model condition does show that contrast occurred in the attractive and not attractive conditions.

Using a no-comparison control group, Thornton and Moore (1993) replicated and extended Cash et al.’s (1983) findings. In comparison to a control group, participants evaluated themselves as more physically attractive after being exposed to non-attractive same-sex stimuli persons and less attractive after exposure to attractive same-sex stimuli persons (Study 3). Men and women displayed the same contrast effects, and social self-esteem differed as a function of condition. Those in the attractive condition rated their social self-esteem lower than those in the non-
attractive and control conditions. Hence, the painful nature of upward comparisons is again displayed in these findings.

**Upward Comparison and Assimilation Effects**

The nineties saw a reversal in the viewpoint that upward comparisons necessarily lead to negative self-evaluations. Researchers began to find that people could also assimilate to upward comparisons, a phenomenon defined by Wheeler and Suls (2007) as “an increase in the comparer’s self-evaluation on a dimension as a result of comparing with someone better on that dimension” (p. 32).

Brewer and Weber (1994) manipulated psychological closeness and provided evidence that upward comparisons could be accompanied by positive evaluations. Participants were randomly assigned to a minority or majority social group and then viewed a video of either an academically outstanding student (upward comparison) or an academically poor student (downward comparison). Participants then rated their intellectual ability relative to other students. When the comparison target was another in-group member, members of the minority group evaluated themselves more positively when they were exposed to the upward target than to the downward target, showing that positive self-evaluations can accrue from upward comparisons.

Other studies began to find that upward comparisons could be inspiring. Lockwood and Kunda (1997) demonstrated that comparing oneself with superstars, an upward comparison, could be a source of inspiration. In a series of experiments, they examined the impact of superstars on self-views. Participants read the description of a superstar whose success was relevant to their future profession (Study 1) and attainable (Study 2). Subsequently, participants rated themselves on adjectives that were positively and negatively related to general career success. In Study 1, future teachers and accountants read an article describing either an outstanding teacher or accountant. Participants rated themselves more positively when the superstar’s success was relevant rather than irrelevant. Due to sample constraints, it was only possible to have a no-comparison control group for the future teachers. These future teachers rated themselves more positively than those in the control group or irrelevant profession condition. In Study 2, first- and fourth-year students read a bogus newspaper article about an all-round fourth-year superstar. A no-comparison control group was included. Compared to first-year controls, first-
year participants who had read about the superstar rated themselves more positively. Conversely, fourth-year participants rated themselves less positively than fourth-year controls, but not significantly so, after being exposed to the superstar. Apparently, the success of the superstar was attainable for the first-year students, but not so for the fourth-year students who were at the same stage as the superstar. Thus the superstar, or the upward comparison, had been a source of self-enhancement and inspiration for these first-year students.

Burleson, Leach, and Harrington (2005) provided evidence that upward comparisons can promote either a sense of inferiority or inspiration depending on an individual’s interpretation of them. These authors investigated the part played by social comparisons in changing the artistic self-concepts of students attending an advanced residential arts programme. The social comparisons that students made were assessed at three time points: before the program began, in the first week of the program, and at the end of the program. Students who interpreted upward comparisons favourably considered them to be inspiring and displayed a corresponding increase in their artistic self-concepts; those who interpreted them unfavourably felt inferior and their artistic self-concepts lessened accordingly. Thus, Burleson et al. concluded that it was the quality of the social comparisons and the way they were interpreted that led to the changes in self-concept, “rather than the mere presence of the talented peers” (p. 120).

Upward Comparison and Task Performance

In other ways, too, upward comparisons have been depicted in a positive light. They have, for example, been shown to enhance performance. Seta (1982) explored the effect of social comparison processes on task performance. Participants were paired with a coactor whose performance level was manipulated to be inferior, identical, or superior to the participant’s. The results showed that a person’s performance on the task was better when paired with a coactor who performed slightly better than she/he did. Interestingly, coactors who performed at identical, inferior, or very superior levels had no effect on the participants’ performances. Similarly, Huguet, Galvaing, Monteil, and Dumas (1999) had their participants perform a Stroop task in the presence of a confederate who worked at a slower, faster or similar rate to the participants (Experiment 2). Participants were made aware of
the speed of the confederate and that the confederate had the same amount of practice as the participants. Response rates to the ink colours of incongruent words (e.g., green printed in blue) and the ink colours of symbols (e.g., + + +) were recorded. Results indicated that participants worked faster (i.e., were less distracted by the Stroop task) when they worked with a confederate whose speed was similar or faster than theirs. Huguet et al. concluded that upward comparisons decreased Stroop interference, thus improving performance (see also Dumas, Huguet, Monteil, & Ayme, 2005; Huguet, Dumas, & Monteil, 2004).

**Upward Comparison and Academic Performance**

In research examining the effects of comparisons on intellectual performance, Stapel and Suls (2004, Study 3) investigated the behaviour of participants on a general knowledge test after they had read information about a very intelligent or unintelligent comparison target. When performance was measured immediately, participants’ intellectual performances were better when they explicitly compared their intelligence with the highly intelligent target rather than with the target of lower intelligence. Similarly, Seaton, Wheeler, and Marsh (2004) exposed their participants to a similarity or difference prime, and to either an academically high-achieving or low-achieving comparison target. Participants then completed the general knowledge sub-test of a widely used intelligence test. Those primed with similarities and who compared themselves with an academically high-achieving target had higher general knowledge scores than those primed with similarities and exposed to a low-achieving target.

In studies outside the laboratory, comparisons with superior others have also displayed similar positive results. In school-based research, upward comparisons have been shown to enhance academic performance. Blanton et al. (1999) asked their participants, first-year high school students, to nominate the student in each of seven academic subjects with whom they preferred to compare their grades. Participants’ school reports were accessed to obtain their grades at three points during the academic year, and used to assess performance and the direction of comparisons. Thus, objective measures of performance and comparison direction were obtained. Students also completed a measure of comparative evaluation, whereby they compared their academic abilities in all seven subjects with those of their classmates.
The authors found that students compared themselves with other students who were slightly outperforming them academically, and that controlling for prior grade, these upward comparisons with selected targets significantly predicted higher grades in subsequent tests. Whereas participants’ own grades predicted comparative evaluation in all seven subjects, choosing to compare themselves with a more able student, however, had no effect on self-evaluations. Blanton et al. concluded that “participants reflected more on their own abilities than on the performances of their targets of comparison when they made their comparative evaluations” (p. 426).

Huguet et al. (2001) replicated Blanton et al.’s (1999) findings using two comparison targets instead of one. Consistent with Blanton et al., they demonstrated that students, in their first year of high school, compared themselves with close friends who were slightly outperforming them academically, and that these upward comparisons were predictive of higher grades. With the exception of one academic subject for comparison choice 2, Huguet et al. also offered evidence that comparing with someone performing better had no effect on one’s self-evaluations. These authors also extended Blanton et al.’s findings in significant ways. They established that students compared themselves with their friends with whom they identified for self-improvement purposes, and that this identification increased with increased academic control, psychological closeness, and importance of academic domains. In recent research, Dumas, Huguet, Monteil, Rastoul, and Nezlek (2005) examined the use of upward comparisons in elementary and middle school students. Whereas seven and eight year olds did not compare upward, students in grades seven and eight (9 and 10 year olds) did for some academic subjects. By the time they reached year nine (12 year olds), students compared upward in all academic subjects, leading these authors to conclude that the tendency to compare oneself with superior others is one that grows stronger over time in the school system.

Section Summary

Downward comparison theory, as proposed by Wills (1981), has been criticised on numerous grounds. However, despite these criticisms, people do compare downwards and when they do they often feel better. Although there may be disagreements concerning the mechanisms involved when people make downward
comparisons, it is clear that they do, as the literature is replete with examples of its use. The research reviewed concerning upward comparisons demonstrates that they can be inspiring and deflating. They can improve performance, bolster self-concept, but can also result in lower self-evaluations. It is little wonder then that writers have described social comparisons, including upward comparisons, as a “double-edged sword” (Major et al., 1991, p. 238; see also Buunk et al., 1990; Diener & Fujita, 1997; Taylor et al., 1990; Testa & Major, 1990).

So, why do people compare themselves with more successful others when in doing so they may lay themselves open to painful experiences? In explaining why individuals compare themselves with superior others, some have suggested that they do so to self-enhance. As Wheeler (1966) stated, “The comparer is attempting to prove to himself that he is almost as good as the very good ones” (p. 30). Collins (1996, 2000) explained the use of upward comparison along similar lines. She posited that comparers who use upward comparisons actually see themselves as possessing the same ability levels and attributes as their objectively better comparison targets. The Nosanchuk and Erickson (1985) study described above is an example of this. Although comparison targets were in fact superior bridge players, participants considered themselves similar in ability to them. Wood (1996) argued that upward comparisons could serve three comparison goals: (1) self-evaluation, obtaining an accurate appraisal of one’s competencies; (2) self-improvement, obtaining information on how to perform better in the future; or (3) self-enhancement, “protecting or enhancing one’s self-esteem” (Wood, 1989, p. 232). Whatever the reason, it is evident that upward comparisons can be both pleasurable and painful (Brickman & Bulman, 1977).

A Theory of Assimilation and Contrast in Social Comparison – The Selective Accessibility Model

So, if a social comparison can result in assimilation, contrast, or both, then how can this be explained? One theory that accounts for these opposing results is the selective accessibility (SA) model (Mussweiler, 2001b). This model has drawn on research from the social cognition tradition to explain the effects of making a social
comparison. Initially formulated to account for the anchoring effect (Strack & Mussweiler, 1997), it has been suggested that the SA model can explain some of the diverse and often contradictory findings in the social comparison literature (Mussweiler, 2001b).

**The Selectivity Hypothesis**

The SA model hinges on two hypotheses, the first being the selectivity hypothesis. In the anchoring paradigm, participants are asked to compare a target to a standard. For example, Mussweiler and Strack (1999, Study 1) administered a questionnaire containing eight pairs of comparative and absolute questions to their participants. Questions varied from estimating the mean temperature in winter in the Antarctic, to Gandhi’s age. To illustrate how these questions were used, the question concerning the River Elbe is herein considered. For the comparative task, half of the participants were asked whether they thought that the River Elbe (target) was longer than 890 kilometres (high standard) and the other half were asked whether they thought that it was shorter than 550 kilometres (low standard). Then participants were asked to answer in absolute terms how long the River Elbe was. The authors found that answers to both the comparative and absolute questions were assimilated towards the standard with which the participants had been primed. If participants had been primed with the high standard, their comparative and absolute judgements were higher than if they had been primed with the low standard. These authors suggested that when solving the comparative task, participants contemplated the probability that the target (River Elbe) and standard (890 or 550 kilometres) were equivalent by selectively looking for information about the target to show that this was the case. When asked to complete the absolute judgement test, participants used the information that was generated in solving the comparative task, and hence made readily available. In applying these findings to social comparison, Mussweiler and Strack (2000a) suggested that when people make a social comparison, they generate knowledge that they are similar to the other person (often called standard-consistent knowledge), by looking for proof that they are equal on the comparison dimension. Thus standard-consistent information is selectively retrieved from the immense self-knowledge that people have, in order to test this similarity premise.
The second hypothesis of the SA model concerns the accessibility of this selectively retrieved information. Mussweiler and Strack (2000a) noted that social cognition research has shown that activating knowledge can affect later evaluations. For example, Srull and Wyer’s (1979, Study 1) participants competed a sentence construction task that was used to prime the concept of hostility. Some participants completed 30 of these items, others 60. Subsequently, some participants continued with the experiment, but others were told to return one hour later or 24 hours later. When the experiment continued, participants read the description of a person, Donald, and were asked to form an impression of him. The description was made deliberately ambiguous in terms of Donald’s hostility. Participants’ ratings of Donald’s hostility increased with the number of times that the concept of hostility had been primed in the participants. At the same time, ratings of Donald’s hostility decreased (although not completely) as the time interval increased between the priming task and the impression formation task (see also Higgins, Rholes, & Jones, 1977). Mussweiler and Strack (2000a) applied this evidence to social comparison and suggested that just generating knowledge about oneself on the comparison dimension makes that knowledge more accessible and so increases the probability that it will be used in later judgements. Moreover, because the information retrieved tends to be standard-consistent, selected because it fits the hypothesis about the comparison dimension, subsequent judgements will tend to be assimilative in nature.

Empirical Research Support for the SA Model

Empirical support for this model has been demonstrated in various studies. For example, Mussweiler and Strack (2000b, Study 1) asked participants to compare themselves to either a moderately athletic or a moderately unathletic celebrity. They were then given a lexical decision task that consisted of neutral words, nonwords, and words related to being athletic or unathletic. In addition, to ensure that self-related knowledge was being made accessible, and not just semantic knowledge per se, the lexical decision task was subliminally preceded either by a word related to the self-concept (I, my, me), or a word not associated with the self (or, and, when). They discovered that comparing themselves with the moderately athletic standard resulted
in participants recognising words related to being athletic faster than words related to being unathletic. When they compared themselves with the unathletic standard, participants recognised words related to being unathletic faster. But, this association only occurred for the trials where the self-concept was subliminally primed. Consequently, it was proposed that the comparison had selectively generated standard-consistent self-knowledge, which was then accessible for subsequent self-evaluations (for further empirical support for this model see also Mussweiler, 2001b; Mussweiler & Strack, 1999, 2000c).

Revision of the SA Model

The SA model implies that assimilation is the typical consequence of making a social comparison. However, Mussweiler (2001b) suggested that while assimilation might be the “default option” (p. 501), this might not always be the case: Under certain circumstances, contrast might be the result. For example, if the standard of comparison is not realistic enough, not applicable or not relevant, then it may not be used, or subsequent judgements could be contrasted away from the standard (Mussweiler & Strack, 2000a). Moreover, contrast could also be the consequence of a social comparison if there was little psychological closeness between the self and the comparison standard, if the achievements of the comparison standard did not seem attainable (Mussweiler, 2001b), or if comparisons are imposed on the individual (Mussweiler, 2003).

In its original form, the SA model did not adequately explain contrast effects. However, in a recent extension to the model, Mussweiler (2001b) attempted to address this issue by suggesting that people test one of two hypotheses when they compare themselves to others. He proposed that when making a social comparison, people either test for similarities or dissimilarities. So, if people look for similarities when they make a comparison, this should result in assimilation and is probably the most common choice. When people look for differences, contrast should be the result. Mussweiler also assumed that regardless of whether people look for similarities or differences, because they have vast amounts of self-knowledge, they would find information consistent with whichever hypothesis that they were testing. Moreover, the standard-consistent knowledge, which has been generated, would then
be easily accessible for subsequent judgements and self-evaluations. This extension to the SA model has been supported empirically (e.g., Mussweiler, 2001b).

Section Summary

The SA model provides a parsimonious, but complete, explanation for the disparity of findings in the social comparison literature regarding assimilation and contrast effects. In addition to being testable, the SA model offers researchers an “integrative framework for assimilation and contrast as a consequence of comparison in general” (Mussweiler, 2003, p. 480).

Generalised Comparison Others

The focus of many of the social comparison studies previously described was a specific individual comparison target (e.g., Aspinwall & Taylor, 1993; Lockwood & Kunda, 1997). There is evidence to suggest, however, that comparisons are not always made with a specific target in mind. Individuals can also compare their achievements with a generalised other (an average mark or the typical person for example; Marsh et al., in review; Suls, Lemos, & Stewart, 2002). Although some authors use the term generalised other (Marsh et al., in review), others have referred to this type of comparison as a comparative evaluation (Blanton et al., 1999; Huguet et al., 2001), or a comparison with the average person (Chambers, Windschitl, & Suls, 2003), typical others (Eiser, Pahl, & Prins, 2001), or with some kind of average or aggregate (Buckingham & Alicke, 2002). Whatever the terminology, no one person is singled out for comparison; rather, this type of comparison refers to comparisons with groups of people, or with the typical or average person or average achievement score on the dimension being assessed.

Historical Conceptualisations of the Generalised Other

The history of the use of the generalised other in psychology can be traced back to William James (1890) who distinguished between two elements of the self: the I
(self-as-knower) and the Me (self-as-known). But, it was Mead (1934), drawing on this distinction, who popularised the term generalised other. Festinger (1954), too, emphasised group processes in his formulation of social comparison theory. He recognised that individuals do not always single out one individual for comparison purposes, but that sometimes comparisons are made at the group level. The theory was influenced by some of Festinger’s early work on informal communication in small groups (Suls & Wheeler, 2000), so it is understandable that comparisons with groups were included. Hence, although awarded little empirical attention, the generalised other is a construct that has a distinguished history.

**Empirical Research Investigating the Generalised Other**

Research that has investigated the role of the generalised other includes a study by Suls and Tesch (1978), who investigated the type of feedback students found most useful in evaluating their academic performance. After receiving their grades from an in-class exam, students were asked to rate how interested they were in obtaining different kinds of information for the purpose of evaluating their performance in the test. Students could obtain nine types of information including the average score, the highest score, the lowest score, and the most frequent score. Participants were most interested in knowing the average score for the exam. For them, the most diagnostic way to assess their performance was to know how their score differed from the average score.

Klein (1997, Study 2) demonstrated that people use generalised others rather than objective standards to judge their ability on a novel task. Klein’s participants were asked to decide which picture, in several pairs of pictures, was aesthetically superior. Participants were told that professional artists had already selected the ones that were the most aesthetically pleasing, so for each pair of pictures there was a “right” answer. They were then told that they had chosen the right picture either 40% or 60% of the time (an absolute or objective standard) and that the average person had scored either 20% above or 20% below the participant’s score (relative or comparative standard). The effect of the absolute or objective score failed to reach significance, but that of the relative score did. Compared to those who scored below average, those who scored above average rated themselves more highly on general aesthetic
ability and picture-judging ability. So, although an objective standard was available, participants chose to use the comparative information derived from an average score.

Marsh et al. (in review, Study 2) examined the effect of a selected upward comparison (based on an explicit subjective measure) and a comparison with a generalised other (based on the class-average grade, an implicit measure) on participants’ math agency. They found that selecting a more able comparison target had a negative effect on math agency, as did the comparison with the generalised other. When both types of comparison were considered in the same model, both effects were significantly negative, with those of class-average being more negative than those of the selected target. In this case, both sources of comparison were used for self-evaluation purposes, but the comparison with the generalised other had more impact.

In addition to providing evidence for the importance of comparisons with the generalised other, Marsh et al.’s (in review) results highlight the effects of implicit and explicit comparisons and draw attention to the use of objective versus subjective measures of comparison direction. Explicit and implicit social comparisons have been shown to produce different outcomes. For example, Stapel and Suls (2004) noted that implicit comparisons resulted in contrast, but explicit comparisons produced either assimilation or contrast effects, depending on certain conditions. In Marsh et al.’s study, the comparison with a generalised other or class-average was implicit and resulted in a contrast effect (the BFLPE), consistent with predictions from Stapel and Suls. Marsh et al.’s subjective measure of upward comparison was explicit but also produced a contrast effect, a result that is not surprising given that Stapel and Suls expected explicit comparisons to produce assimilation only “under particular boundary conditions” (p. 873). While consistent with a host of other studies showing that upward comparisons can be painful (e.g., Morse & Gergen, 1970; Thornton & Moore, 1993), this contrast effect is contrary to more recent work where objective indicators of upward comparison have produced assimilation effects (Blanton et al., 1999; Huguet et al., 2001). These objective measures are not akin to objective rating scales, whereby it has been shown that objective and subjective rating scales produce different outcomes (Biernat, Manis, & Kobrynowicz, 1997; Manis, Biernat, & Nelson, 1991; Mussweiler & Strack, 2000b). Rather students’
grades are used as objective indicators of comparison direction in these studies. Thus, it is possible that the differences found between the use of objective indicators of comparison direction (Blanton et al., 1999; Huguet et al., 2001) and the subjective rating scale of Marsh et al. are attributable to the different methodologies employed.

Section Summary

Although the generalised other has a distinguished history, it has received little empirical attention. There is, however, empirical evidence to suggest that individuals do use generalised others for self-evaluation purposes. Interestingly though, as Buckingham and Alicke (2002) noted, social comparison theories have yet to clarify the relative importance of specific and generalised comparisons in evaluating one’s competencies.

Comparisons Forced by the Environment

The Distinction Between Selected and Forced Comparisons

Wood (1989) noted, “the literature has overemphasized target selection” (p. 234). While it is true that the selection of comparison targets has been of major interest, many laboratory experiments have forced comparisons on participants. The participants in Morse and Gergen’s (1970) study were placed in a situation where they were forced to compare with Mr. Clean or Mr. Dirty. Participants who view photographs (e.g., Cash et al., 1993) or read a scenario about another person (e.g., Lockwood & Kunda, 1997) are having a comparison forced on them, albeit in the artificial atmosphere of a social science laboratory. These paradigms are not concerned with whom one chooses for comparison; rather they provide a comparison target and investigate its effect on the individual. In doing so, they attempt to simulate real life. In reality, individuals may not always consciously select comparison targets; rather the environment may force comparisons upon individuals. Forced comparisons can occur from discovering that one’s neighbour has acquired an expensive new sports car, to hearing that a co-worker has just received a coveted prestigious promotion. So, unlike selected comparisons where individuals choose
their comparison target, forced comparisons are imposed by the environment and come involuntarily without being sought or chosen.

The differentiation between chosen and forced comparisons was implied by Brickman and Bulman (1977), although they did not overtly distinguish the two. For instance, they described the situation of a class reunion that people can choose to go to, but where social comparison is the purpose of the event. When people attend a class reunion, although they have chosen to attend, the comparisons are forced on them – they have little option but to compare their success (or failure) with that of their former schoolmates. These may not be comparisons that people would have chosen to make, but comparisons that are thrust on them by virtue of attending such an event.

Various writers (e.g., Suls, 1986) have emphasised this distinction between forced and selected social comparisons. Goethals (1986) traced the history of social comparison theory and discussed new directions for the discipline, including the need to recognise the existence of forced comparisons (whose existence was acknowledged by Festinger, 1954) and to investigate the processes involved in both selected and forced comparisons.

**Research Evidence for Forced Comparisons**

Evidence for the forced social comparison paradigm comes from Marsh and Peart (1988) who examined the effects of two different types of fitness programs on the fitness levels and physical self-concepts of high school girls. Pre-test physical fitness scores were used to divide the students into quartiles and then students from each quartile were randomly assigned to three experimental groups: a control group who played an unstructured game of volleyball, and two experimental groups who participated in an aerobics training program. Of the two experimental groups, one received social comparison feedback (e.g., “Look how many Mary did. She’s miles ahead of the rest of you – come on”, p. 398), while the other received cooperative improvement feedback (e.g., “You all handled that exercise very well”, p. 397). Compared to the control group, the fitness of both the social comparison feedback group and the cooperative improvement feedback group improved. However, there
was a difference in the physical self-concepts of these groups: The physical self-concept of those who received cooperative improvement feedback increased from pre-test levels, but declined for those who received social comparison feedback. Scores for the control group were midway between those of the two experimental groups, but did not differ significantly from either. Results indicated that after the intervention program, those in the cooperative improvement group had higher physical self-concepts than those in the social comparison group. Furthermore, in the social comparison group the students would have been compelled to compare with other group members, as the comparison was made particularly salient in that experimental group.

Diener and Fujita (1997) reviewed evidence for the effects on subjective well-being of forced and selective (although the term they used was coping) approaches to social comparison. Although initial evidence regarding the effect of forced comparisons was positive, subsequent studies demonstrated that forced social comparisons had no long-term effects on subjective well-being. Additionally, they concluded that the use of selected comparisons could overcome the effects of forced comparisons. However, they noted that Marsh’s findings on the BFLPE (see Chapter 2 on the BFLPE) were one area where imposed social comparisons could influence subjective well-being. These authors suggested that the reason that Marsh’s findings were at odds with other findings on comparisons imposed by the environment was that the school, the basis of Marsh’s research, was a “total environment” (p. 350) that provides a standard frame of reference for all. Diener and Fujita suggested that individuals might normally obtain social feedback from many different sources, but that in the total environment of the school the information that an individual could receive would be greatly narrowed. As they speculated, “schools more nearly approximate “total environments” (strongly controlling the information that the individual receives about the distribution of academic abilities, and also emphasizing feedback that has primarily a comparative meaning) than the environments examined in other comparison studies” (p. 351). In schools, the frame of reference that students have to rely on consists only of other students in the school and the information that teachers provide about grades and their distribution. Thus, students are left with little option but to compare themselves with other students in their school to evaluate their performance. In this case, even adaptive selective comparisons may not be able to
overcome those forced on students by the total environment of the school. Students may not have a choice about whether or not to use the comparison information imposed on them by the school environment – it may be the most diagnostic tool they have. Hence in certain circumstances, such as Marsh’s BFLPE, “the imposed environment can influence a person’s self-concept and probably his satisfaction as well” (Diener & Fujita, 1997, p. 351).

In their study of downward comparison in everyday life, Wood et al. (2000) asked their participants to keep a diary of comparisons they made over a three-week period. These authors were interested in their participants’ motivations for comparisons. As such, one of their diary questions pertained to whether the comparison was voluntary or forced. Although 66% of reported comparisons were motivated for self-evaluative, self-enhancement or self-improvement reasons, about one third were unintended, that is, comparisons that were involuntary and encountered in the course of day-to-day living. Wood et al. commented that this was not a surprising finding considering that comparisons are an “almost inevitable element of social interaction” (Brickman & Bulman, 1977, p. 150, as cited by Wood et al., 2000, p. 571).

The Burleson et al. (2005) study reported above is also an example of forced social comparisons. Recall that the students in that study were participating in a month-long residential advanced arts program. In this situation, the comparisons could be regarded as being forced, as students would have had little option but to compare themselves with other highly talented students.

**Section Summary**

Although he acknowledged that social comparisons could be imposed by the environment, Festinger’s (1954) original theory emphasised that the individual was active in the comparison process and since then much of the social comparison literature has emphasised target selection (Wood, 1989). However, researchers have acknowledged, and demonstrated, that social comparisons are not always sought but can also be forced on individuals by the environment in which they find themselves. As Wood noted, the environment is not an “inactive backdrop” but rather is able to
influence one’s self-evaluations (p. 233). In support of this argument, Wood cited the BFLPE findings of Marsh and others, noting that they “would not be included in the traditional social comparison literature” but that they “portray a very active role for the social environment” (p. 233). This relation between Marsh’s BFLPE and social comparison will be explored in the following section.

Social Comparison and the BFLPE

Forced Comparisons with the Generalised Other

The theory underpinning Marsh’s BFLPE model (see Chapter 2 for details of the BFLPE) encompasses both generalised and forced social comparisons, neither of which has been explored fully by social comparison researchers – perhaps one of the reasons that the model has not had more impact on the mainstream social comparison literature. Marsh and his colleagues have proposed that the theoretical basis of the BFLPE lies in social comparison theory (e.g., Marsh et al., 1995; Marsh & Hau, 2003; Marsh et al., 2000; Marsh et al., in review). In explaining the development of the BFLPE, they have suggested that students form their academic self-concepts partly by comparing their academic performance with that of their peers. In academically selective settings these peers are other highly intelligent students and so comparisons can lead to students feeling less able with a resulting lowering of academic self-concepts. In non-selective academic environments, where there is a much wider ability range with which to compare, students are able to compare themselves with less intelligent students and so academic self-concepts tend to be higher. So, for example, an average student in a high-ability school may be performing academically below the average achievement level for that school. That student may then compare his/her academic ability with other students in that school (i.e., a generalised other) and this may result in a lower academic self-concept. However, if this same student were attending an average-achievement school he/she may well perform above the average for that school, with a resulting increase in academic self-concept (Marsh & Hau, 2003). In making these comparisons, the environment imposes a comparison on students as they can only use the frame of reference available to them to evaluate their ability – other students in their school (i.e., a forced comparison).
The BFLPE and Low-Ability Students

The validity of the proposition that social comparison underlies the BFLPE would be strengthened were low-ability students in regular classes also to display BFLPEs. Marsh (2005) noted that although most BFLPE research has concentrated on the negative effects of attending high-ability classes or schools on high-ability students, the theory also has implications for low-ability students. If low-ability students are educated with other low-ability students in streamed classes, then they should have higher academic self-concepts than low-ability students placed in regular mixed ability classes. This is because in streamed classes comparison targets are students of similar ability, but in regular classes, comparison targets for low-ability students would presumably be more able on average (Marsh, 2005; Marsh & Craven, 2002). This hypothesis was examined in a longitudinal study of the academic self-concepts of academically disadvantaged students in regular mixed-ability classes, and of those who moved to special support units (Tracey, Marsh, & Craven, 2003; see also Marsh, Tracey, et al., 2006). Thirty-nine intellectually disadvantaged students were assessed at three time points. At T1 all were enrolled in regular mixed ability classes. At T2, six months later, 21 had moved into special support units. The last assessment was made 10 academic months after students had been placed in their respective classes. Over time there was a relative decrease in the academic self-concepts of students in the regular mixed ability classes, but an increase in the academic self-concepts of the students placed in the special support units. The intellectually disadvantaged students in regular classes had been exposed to relatively more able students, unlike their counterparts in the support units, with a resulting drop in their academic self-concepts. Although other explanations are feasible, such as the students in the support units were given special support, it is also possible that comparisons with the more able students had been at least partly responsible for the decline.
**Section Summary**

Thus, the theory underpinning the BFLPE rests on two social comparison premises, i.e., that comparisons are made with a generalised other, in addition to comparisons with individuals; and that comparisons with the generalised other are forced on students by the school environment. The use of selected individual comparison targets and forced generalised targets is not mutually exclusive. Students may have different motivations in using each of them and their subsequent effects may not always be similar. However, to date, the differential effects of selected and forced comparison targets have not been well documented (but see Marsh et al., in review).

**Implications for the Present Investigation**

*The Blanton et al. (1999) and Huguet et al. (2001) Studies*

Wheeler and Suls (2005) were the first to note the contradiction between the BFLPE results of Marsh and his colleagues and the social comparison findings of Blanton et al. (1999) and Huguet et al. (2001). The Blanton and Huguet studies imply that upward comparisons improve academic performance, but the BFLPE findings suggest otherwise. The BFLPE studies of Marsh and his colleagues imply that comparisons deflate self-evaluations of academic ability and lower academic self-concepts have been associated with lower achievement (e.g., Chapman et al., 2000; Hansford & Hattie, 1982). Furthermore, related research has demonstrated that academic self-concept and achievement have a reciprocal relation such that improvements in academic self-concept lead to higher achievement, but improvements in achievement also lead to higher academic self-concept (REM, see Guay et al., 2003; Marsh & Craven, 2005, 2006; Marsh et al., 2002; Marsh & Yeung, 1997a; Valentine & DuBois, 2005). Herein then lies an apparent discrepancy between these two strands of psychological research. Whereas social comparison studies have demonstrated that selected upward comparisons enhance performance, BFLPE results suggest that comparisons based on the generalised other deflate self-evaluations of academic ability, and lower academic self-concepts have been associated with lower levels of achievement. The next logical step in the progression...
of this field of intellectual inquiry is therefore an attempt to reconcile these two apparently inconsistent sets of results, each based on well established research findings that have been replicated in subsequent studies.

**The Theoretical Basis of the BFLPE – Social Comparison Theory**

Researchers who study the BFLPE propose that its theoretical basis lies in social comparison theory (Marsh et al., 1995; Marsh & Hau, 2003; Marsh et al., 2000), but to date this causal link has not been investigated fully in relation to the implicit or explicit comparison processes that individuals actually use. BFLPE theory posits that comparisons are made with a generalised other forced by the environment, rather than being made with a selected individual. To date, very little research has been devoted to exploring the differential effects of these two types of comparisons. Hence, an important issue for BFLPE and social comparison researchers alike is to examine differences that may exist between selecting an individual for comparison and having a generalised comparison imposed, in an attempt to explicate the effects of each.

**The Present Investigation**

The overarching purpose of the present investigation is to address some of the issues discussed in Chapter 2 and Chapter 3 above by extending current BFLPE theory and research with the intent of elucidating potentially potent processes and constructs to inform new solutions for enabling high-ability students to reach their full potential. More specifically, the broad aims of the present investigation are to capitalise on cutting-edge advances in BFLPE theory and research and strong statistical analyses to: (a) test the external validity of the BFLPE across 41 countries in order to ascertain whether support can be identified for the universality of this theory; (b) test whether the BFLPE is evident in economically developing countries and collectivist countries to ascertain whether the BFLPE transcends cultural and economic barriers; (c) investigate potential moderators of the BFLPE to identify constructs (such as individual ability, socio-economic status, individual differences
in learning, and individual perceptions of the learning environment) that may moderate the adverse effects of the BFLPE; (d) elucidate whether upward social comparisons moderate or co-exist with the BFLPE to resolve a conflict in the literature; (e) critically analyse the relation between social comparison processes and the BFLPE to further inform theory; and (f) test whether the BFLPE varies as a function of ability to clarify the impact of the BFLPE on different levels of student ability. Three synergistic studies were designed to address these aims.

Study 1 (see Chapter 4), the largest BFLPE study undertaken to date, was designed to: (a) test the external validity of the BFLPE in 41 countries; (b) test if the BFLPE was present across collectivist versus individualist countries and across economically developed versus economically developing countries; and (c) investigate potential moderators of the BFLPE such as individual student characteristics, socio-economic influences, and individual perceptions of the learning environment. The sample comprised 265,180 15-year-old students from 41 countries who participated in the OECD’s PISA study in 2003 (see Organisation for Economic Cooperation and Development, 2005a, 2005b). Multilevel modelling was used to ascertain whether the BFLPE was evident in the total sample of 41 countries, each country individually, and in economically developing countries and collectivist countries as well as in developed countries and individualist countries. In addition, 18 constructs were investigated in this study to determine if any aggravated or alleviated the negative effects of the BFLPE. Hence, this study was designed to extend BFLPE theory and research in three important ways by: (1) providing a rare large-scale BFLPE study based on a sample of 41 countries; (2) implementing a diverse cross-cultural study of the BFLPE to enable tests of its universality across different types of cultures and nations; and (3) assessing a vast array of potential BFLPE moderators.

The second and third studies explored the relation between the BFLPE and social comparison processes, while also examining potential BFLPE moderators. Study 2 (see Chapter 5) tested whether selected upward social comparisons, which have been shown to enhance performance, moderate or co-exist with the BFLPE. This was accomplished by undertaking a further analysis of previously published social comparison data (Blanton et al., 1999; Huguet et al., 2001) from a BFLPE
perspective. Hence, Study 2 was designed to: (a) replicate the original results of the two social comparison studies; (b) test for the presence of the BFLPE in the same data that demonstrated that selected upward social comparisons enhanced performance; and (c) to ascertain whether upward social comparisons moderate or co-exist with the BFLPE, in order to elucidate the relation between the BFLPE and social comparison theory. Participants were 876 Dutch students and 1,156 French students in their first year of high school, who had taken part in the original social comparison studies. Not only did this study bring together two theoretical perspectives in the context of a single study, but also by re-analysing previously published data using more powerful statistical methodology Study 2 made a significant contribution.

Study 3, described in Chapter 6, consisted of a critical analysis specifically intended to test the relation between selected social comparisons and the BFLPE. A total of 2,015 French students completed standardised tests and a questionnaire that assessed self-concept, comparison direction, and motivational orientation. Hence, this study was one of the first to have as its primary concern an investigation of social comparison processes as they relate to the BFLPE. Thus, Study 3 was designed to: (a) further elucidate the relation between selected social comparisons and the BFLPE by addressing previous limitations through the use of psychometrically sound instrumentation and standardised achievement tests; and (b) test whether individual differences in achievement goal orientation moderate the BFLPE in order to identify characteristics of individuals that may alleviate or aggravate the negative effects of the BFLPE. Additionally, all three studies were designed to investigate the nature of the relation between the BFLPE and individual student ability levels in order to explicate whether the BFLPE varies as a function of individual ability.

The specific aims, hypotheses and their rationales, research design, methodology, and results for each of these three studies are contained within their relevant chapters. These chapters (4, 5, and 6) also contain a discussion of findings that are pertinent to the particular study. The strengths and limitations of each study, as well as implications for theory, future research, and practice are also presented within each chapter.
Summary

This chapter has reviewed the relevant theoretical and empirical literature pertaining to social comparison processes and has outlined the relation between social comparison processes and the BFLPE. Implications arising from these discussions, which were of significance to the present investigation, have also been emphasised. The chapter ended by summarising the three studies that comprise the present investigation. In the next chapter the aims, results, and discussion of results for Study 1 are presented.
CHAPTER 4

STUDY 1: TESTS OF THE EXTERNAL VALIDITY OF THE BIG-FISH-LITTLE-POND EFFECT AND A SEARCH FOR MODERATORS

Introduction

As discussed in Chapter 2, whilst the BFLPE has been researched extensively in developed countries and individualist countries, with a few exceptions (e.g., Hong Kong, Israel) collectivist and economically developing countries have for the most part been neglected in BFLPE research. The most impressive study to include developing countries and collectivist countries was a cross-national study of 26 nations by Marsh and Hau (2003); see Chapter 2. Study 1 of the present investigation was designed to replicate and extend their study by: (a) expanding the sample to 41 nations; (b) including 16 collectivist and 14 developing countries; (c) testing if the BFLPE differentially operates in collectivist and individually-orientated nations; (d) highlighting the extreme domain specificity of the BFLPE; and (e) explicating moderators of the BFLPE. If the BFLPE were to span cultures, this would indeed be testament to its universality. Thus, in broad terms Study 1 endeavoured to establish the external validity and universal applicability of the BFLPE, by demonstrating its existence in collectivist countries and the developing world. In doing so, Study 1 represents the largest cross-cultural study ever undertaken in BFLPE research, in terms of the sheer size of the sample and of the number and diversity of the countries investigated. Additionally, researchers have advocated that the identification of individual differences between students that moderate the negative effects of the BFLPE would be a valuable step towards developing policies aimed at maximising the benefits and minimising the negative effects of attending academically selective classes and schools. To date research investigating these factors has had limited success (see Chapter 2). Thus, Study 1 also aimed to advance knowledge of the BFLPE and add to the scientific literature in this field by identifying and examining potential moderators of the BFLPE.

The primary objective of the current chapter is to describe Study 1 by presenting: (a) the nature of the problem under investigation, (b) background to the problem, (c)
overarching aims, (d) specific hypotheses to be tested, (e) research questions to be addressed, (f) rationale for hypotheses and research questions, (g) methods employed, and (h) results of the investigation in the context of the hypotheses and research questions posed. A discussion of the results and the implications thereof for theory, research, and practice is also presented.

Statement of the Problem

Is the BFLPE present across a sample of 41 countries? Is the BFLPE only found in developed countries and in individualist countries, or does it transcend economic and cultural barriers and extend to collectivist countries and the developing world? Does the BFLPE vary as a function of individual ability? Are there any individual difference constructs that moderate the negative effects of the BFLPE?

Background

External Validity of the BFLPE

As noted in Chapter 2, for over 20 years Marsh and his colleagues have demonstrated the existence of the BFLPE – that students in high-ability classes and schools have lower academic self-concepts than similar ability students in low-ability environments. In doing so, these authors have documented the negative effects of attending high-ability classes and schools. Throughout these years there has been considerable cross-national support for the BFLPE in the USA, Australia, Europe, the Middle East and Hong Kong (e.g., Craven et al., 2000; Marsh, 1987, 1991, 2004; Marsh et al., 1995; Marsh et al., 2000; Marsh & Parker, 1984; Mulkey et al., 2005; Zeidner & Schleyer, 1998; see Chapter 2). Additionally, Marsh and Hau (2003) found evidence of the BFLPE in the largest cross-national BFLPE study undertaken to date, in which 26 countries displayed a BFLPE. However, very few of these countries could be considered to be collectivist or developing countries (see Chapter 2). Hence, the vast majority of the countries examined so far have been developed countries and individualist countries.
To establish the external validity, universality, and generalisability of the BFLPE, it is necessary to demonstrate its existence in countries differing in cultural values and economic development. As Schwartz and Bilsky (1990) note, “Theories that aspire to universality, … must be tested in numerous, culturally diverse samples” (p. 878). Theories that span cultures remind us of our common humanity, and compared to theories that are culturally dependent, they are regarded as more externally valid and more generally applicable. To find that the BFLPE transcends economic and cultural barriers and extends to the more collectivist and less economically developed countries of the world would indeed be testament to its universality.

To explore the external validity of the BFLPE, Study 1 utilised the PISA database developed by the OECD in 2003 (see Organisation for Economic Cooperation and Development, 2005a, 2005b). PISA is an internationally standardised assessment conducted by the OECD that takes place every three years and is administered to 15-year-old students in schools. The first assessment took place in 2000, the next in 2003, and the most recent one was undertaken in 2006. The assessments typically focus on math, literacy, and science, and assess, not just achievement per se, but to what extent students, who have nearly completed their compulsory education, have gained the knowledge and skills necessary to be active participants in society. Forty-three countries took part in 2000, 41 in 2003, and it is expected that 58 countries will have participated in 2006. All participating countries cooperated to develop the PISA instruments. Apart from achievement data, PISA also contains a vast amount of additional information. The database contains information about students’ families, home backgrounds, and socio-economic status. Students also report their learning habits, how they feel about their achievements, their perceptions of their schools, and their computer knowledge (Organisation for Economic Cooperation and Development, 2005b).

Marsh and Hau (2003) conducted their cross-national BFLPE analyses using the PISA 2000 database (see Organisation for Economic Cooperation and Development, 2001a, 2001b). As noted in Chapter 2, only five countries in their sample could be regarded as collectivist countries and only six as developing countries. In 2003, the OECD gathered data for a second PISA database, adding new countries. Of these, 16 could be regarded as collectivist (see Hofstede & Hofstede, 2005, for country
classifications), and 14 are considered by the World Bank to be developing (see World Bank, 2007, for economic classifications). The availability of these new data, encompassing as they do numerous culturally and economically diverse countries, therefore allows the validity and generalisability of the BFLPE to be further tested by ascertaining whether its effects extend to developing and collectivist countries. It is anticipated that this comprehensive examination of the BFLPE will ascertain whether the BFLPE is only found in developed countries and individualist countries, or whether it is a truly universal phenomenon.

A Search for Moderators

That students in high-ability schools should have lower academic self-concepts than their counterparts in low-ability schools (the BFLPE) is a cause for concern, as low academic self-concept has been associated with a number of undesirable educational outcomes such as poorer occupational or educational aspirations, lower grade-point-averages, and the selection of less demanding courses (Marsh, 1991; Marsh & Yeung, 1997b; see Chapter 2). Furthermore, the REM described in Chapter 2 (e.g., Guay et al., 2003; Marsh & Yeung, 1997a) suggests that academic self-concept and academic achievement are reciprocally related, such that high academic self-concepts are related to high academic achievement and vice versa. If, as research has shown, high-ability students attending high-ability classes and schools have lower academic self-concepts then it follows that they may not be reaching their full academic potential. To overcome the BFLPE and so allow these students to achieve their best, it is necessary to firstly identify whether there are any factors that can attenuate the BFLPE. Once identified, these factors could be used to assist in developing educational policies aimed at maximising the benefits, and limiting the negative effects, of attending academically selective environments. Although previous research has begun this search for factors that may moderate the BFLPE, success has been limited (e.g., Marsh & Hau, 2003; Marsh, 1984, 1987, 1991). Study 1 continued this search by identifying 20 constructs, within four broad areas, described in Chapter 2, that have the potential to moderate the BFLPE. These areas are: individual ability, socio-economic status, individual differences in learning, and individual perceptions of the learning environment. As few of these constructs have
been fully investigated previously in relation to the BFLPE, Study 1 was designed to fill this gap in the literature by examining these moderators in the context of the BFLPE.

This investigation can be undertaken using the PISA 2003 database (see Organisation for Economic Cooperation and Development, 2005a, 2005b) as this database contains background information that students completed about themselves, their homes, and their schools. For example, students were asked about their reading habits, learning strategies, motivation, their parents’ employment, and their home environment. Students also completed questions assessing classroom atmosphere, relationships with teachers, and the character of the school. This information offers the potential to elucidate individual differences in students’ backgrounds, in their learning habits, and in their perceptions of their learning environments that may predict which students will benefit from attending a high-ability school and which students will not. Thus, by examining individual learning differences and differences in home and school environments, it is hoped that important individual differences will be identified that may serve to alleviate the effects of the BFLPE.

Although there is a solid theoretical basis underpinning the constructs identified as potential BFLPE moderators, by its very nature this search for moderators is a “fishing expedition”. Fortunately, the large sample size available to the present investigation meant that the typical problem of fishing expeditions – capitalising on chance – was virtually eliminated. Also, it seemed appropriate to embark on a fishing expedition when investigating the big-fish-little-pond effect!

**Section Summary**

Very few studies have examined the BFLPE in developing or collectivist countries and there has been little success in discovering factors that have the potential to alleviate the negative effects of the BFLPE. Study 1 sought to overcome these shortcomings by determining whether the BFLPE generalised to collectivist countries and developing countries, and by identifying potential moderating constructs.
Aims

The overarching goal of Study 1 was to empirically test the external validity and generalisability of the BFLPE in 41 countries and to ascertain whether its effects extended to economically developing and collectivist countries. More specifically Study 1 aimed to:

1 Test the external validity and generalisability of the BFLPE by investigating (a) whether the BFLPE is evident in economically developed countries, individualist countries, economically developing countries and collectivist countries, both cross-nationally and on an individual country basis; (b) the extent to which the BFLPE varies between countries; and (c) whether the size of the BFLPE varies as a function of culture or economic development;
2 Test whether the size of the BFLPE varies across different levels of individual student ability in order to explicate whether the BFLPE varies as a function of individual ability; and
3 Elucidate whether particular individual student characteristics (self-regulated learning, motivation, self-efficacy, math anxiety, preference for learning environment), socio-economic influences, and individual perceptions of the learning environment moderate the BFLPE in order to identify characteristics of individuals that may moderate the BFLPE.

Statement of the Hypotheses and Research Questions

When past research and theory were unable to provide guidance for formulating clear predictions, research questions were devised; otherwise, hypotheses were formulated. Hypotheses and research questions are grouped according to the aim from which they were derived. The numbers pertaining to the hypotheses and research questions correspond to the number of the aim that they were developed to satisfy. Hence, for example, Hypothesis 1.1 is the first hypothesis pertaining to the first aim and Research Question 1.2 is the second research question based on the first aim.
The External Validity and Generalisability of the BFLPE

**Hypothesis 1.1: The Relation Between Individual Math Ability and Math Self-Concept in a Cross-National Sample**

Across the cross-national sample consisting of 41 individualist, collectivist, developed, and developing countries, there will be a statistically significant positive relation between individual math ability (both linear and quadratic components, as measured by achievement in math) and math self-concept (measured by a math self-concept scale).

**Hypothesis 1.2: The Relation Between School-Average Math Ability and Math Self-Concept in a Cross-National Sample**

Controlling for individual math ability, there will be a statistically significant negative relation between school-average math ability (as measured by the mean math achievement level for each school) and math self-concept as predicted by the BFLPE (see Figure 2.3) across the entire cross-national sample of 41 individualist, collectivist, developed, and developing countries.

**Hypothesis 1.3: Country Variation in the BFLPE**

Across the cross-national sample of 41 individualist, collectivist, developed, and developing countries, there will be a statistically significant, but small, variation in the relation between school-average math ability and math self-concept.

**Hypothesis 1.4: The Relation Between Individual Math Ability and Math Self-Concept in Individual Countries**

In each of the 41 individual countries of the cross-national sample of individualist, collectivist, developed, and developing countries, there will be a statistically significant positive relation between individual math ability (both linear and quadratic components, as measured by math achievement) and math self-concept.
Hypothesis 1.5: The Relation Between School-Average Math Ability and Math Self-Concept in Individual Countries

In each of the 41 individual countries of the cross-national sample of individualist, collectivist, developed, and developing countries, controlling for individual math ability, there will be a statistically significant negative relation between school-average math ability (as measured by the mean math achievement level for each school) and math self-concept (the BFLPE; see Figure 2.3). Although it is expected that the direction will be the same in individual countries, the size of the relationship will be different across countries.

Research Question 1.1: Culture as a Moderator of the BFLPE

Can the cultural orientation of a country moderate the BFLPE? In the cross-national sample, what will be the main effect of cultural orientation (as measured by an Individualism Index scale, a scale that rates countries from individualist to collectivist, developed by Hofstede & Hofstede, 2005) on math self-concept? What relation will exist between the school-average math ability X cultural orientation interaction and math self-concept?

Research Question 1.2: Economic Development as a Moderator of the BFLPE

Can the economic development of a country moderate the BFLPE? In the cross-national sample, what will be the main effect of the economic development of a country (as measured by the World Bank economic classifications scale, World Bank, 2007), on math self-concept? What relation will exist between the school-average math ability X stage of economic development interaction and math self-concept?

Individual Ability and the BFLPE

Research Question 2.1: Individual Math Ability as a Moderator of the BFLPE

Does individual math ability (linear and quadratic components) moderate the BFLPE? In the cross-national sample of 41 countries, what is the relation between the individual math ability (both linear and quadratic components) X school-average math ability interaction and math self-concept?
**Individual Student Characteristics and the BFLPE: SES, Individual Differences in Learning, and Perceptions of the Learning Environment**

**Hypothesis 3.1: Relation Between Individual Socio-Economic Status (SES) and Math Self-Concept**

In the cross-national sample, there will be a statistically significant positive relation between each of the individual SES variables (highest in occupation, highest in education, home education resources, cultural possessions, and economic, social and cultural status) and math self-concept.

**Research Question 3.1: Individual SES as a Moderator of the BFLPE**

Does SES moderate the BFLPE? In the cross-national sample, what is the relation between each of the individual SES variables (highest in occupation, highest in education, home education resources, cultural possessions, and economic, social and cultural status) X school-average math ability interactions and math self-concept?

**Hypothesis 3.2: Relation Between Self-Regulated Learning Strategies and Math Self-Concept**

In the cross-national sample, there will be a statistically significant positive relation between each of the self-regulated learning strategy variables (elaboration, memorisation, and control strategies) and math self-concept.

**Research Question 3.2: Self-Regulated Learning Strategies as Moderators of the BFLPE**

Do self-regulated learning strategies moderate the BFLPE? In the cross-national sample, what is the relation between each of the interactions of the self-regulated learning strategies (elaboration, memorisation, and control strategies) X school-average math ability and math self-concept?

**Hypothesis 3.3: Relation Between Motivation and Math Self-Concept**

In the cross-national sample, there will be a statistically significant positive relation between both intrinsic and extrinsic motivation and math self-concept.
Research Question 3.3: Motivation as a Moderator of the BFLPE

Does motivation moderate the BFLPE? In the cross-national sample, what is the relation between the motivation (both intrinsic and extrinsic) X school-average math ability interactions and math self-concept?

Hypothesis 3.4: Relation Between Math Self-Efficacy and Math Self-Concept

In the cross-national sample, there will be a statistically significant positive relation between math self-efficacy and math self-concept.

Research Question 3.4: Math Self-Efficacy as a Moderator of the BFLPE

Does math self-efficacy moderate the BFLPE? In the cross-national sample, what is the relation between the math self-efficacy X school-average math ability interaction and math self-concept?

Hypothesis 3.5: Relation Between Math Anxiety and Math Self-Concept

In the cross-national sample, there will be a statistically significant negative relation between math anxiety and math self-concept.

Research Question 3.5: Math Anxiety as a Moderator of the BFLPE

Is the BFLPE moderated by math anxiety? In the cross-national sample, what is the relation between the math anxiety X school-average math ability interaction and math self-concept?

Hypothesis 3.6: Relation Between Preference for Learning Environment and Math Self-Concept

In the cross-national sample, there will be a statistically significant positive relation between both types of preferences for learning environment (competitive and cooperative) and math self-concept.

Research Question 3.6: Preference for Learning Environment as a Moderator of the BFLPE

Does preference for a learning environment moderate the BFLPE? In the cross-national sample, what is the relation between each of the preferences for learning
environment (competitive and cooperative) \( \times \) school-average math ability interactions and math self-concept?

**Hypothesis 3.7: Relation Between Individual Perceptions of the Learning Environment and Math Self-Concept**

In the cross-national sample, there will be a statistically significant positive relation between individual perceptions of the learning environment (perceptions of student-teacher relations, a sense of belonging to the school, and attitudes to school) and math self-concept.

**Research Question 3.7: Individual Perceptions of the Learning Environment as Moderators of the BFLPE**

Is the BFLPE moderated by individual perceptions of the learning environment? In the cross-national sample, what is the relation between each of the individual perceptions of the learning environment (perceptions of student-teacher relations, a sense of belonging to the school, and attitudes to school) \( \times \) school-average math ability interactions and math self-concept?

**Section Summary**

Grouped according to the aims they were developed to satisfy, the hypotheses and research questions to be examined in the current study were presented in this section. The following section reviews the rationales on which these hypotheses and research questions were based. As with the hypotheses and research questions, the rationales are grouped according to the aims of the current study.
Rationale for Hypotheses and Research Questions

The External Validity and Generalisability of the BFLPE

Rationale for Hypothesis 1.1: Relation Between Individual Math Ability and Math Self-Concept in a Cross-National Sample

The cross-national sample of the current study contains 41 countries differing in cultural orientation and economic prosperity. Research in developed countries and individualist nations has consistently demonstrated a statistically significant positive relation between individual ability (linear) and various facets of academic self-concept (e.g., Marsh, 1991; Marsh & Parker, 1984; Marsh et al., in press). Although very few studies have investigated these relations with collectivist or economically developing countries, those that have, have found that individual ability (linear) has a positive effect on academic self-concept (e.g., Marsh et al., 2000; Zeidner & Schleyer, 1998). In their cross-national study of 26 countries, Marsh and Hau (2003) demonstrated a significantly positive relation between individual ability (both linear and quadratic) and academic self-concept. Thus, it seems appropriate to predict that similar relations will be found between individual ability (linear and quadratic) and math self-concept in the cross-national sample of the current study.

Rationale for Hypothesis 1.2: Relation Between School-Average Math Ability and Math Self-Concept in a Cross-National Sample

Individualist countries, collectivist countries, economically developed countries, and economically developing countries comprise the 41 countries of the cross-national sample. Considerable empirical research in developed countries and individualist countries has demonstrated that, controlling for individual ability, class- and school-average ability have a statistically significant negative relationship with academic self-concept (e.g., Craven et al., 2000; Marsh, 2005; Marsh & Craven, 2002; Marsh et al., 2001; Marsh & Rowe, 1996; Marsh et al., in press; Mulkey et al., 2005). Of the little research that has been conducted in collectivist countries and developing countries, results have been similar (e.g., Marsh et al., 2000; Zeidner & Schleyer, 1998). For example, in their cross-national study of 26 countries, Marsh and Hau (2003) noted that school-average ability had a significantly negative effect on academic self-concept. On this basis it seems reasonable to predict that there will
be a negative relation between school-average math ability and math self-concept in the cross-national sample of the present investigation.

**Rationale for Hypothesis 1.3: Country Variation in the BFLPE**

In their investigation of the BFLPE in 26 countries, Marsh and Hau (2003) demonstrated that there was a statistically significant, but small, variation in the effect of school-average ability on academic self-concept between countries. The current study incorporated many more countries (41) and so a similar statistically significant variation is predicted in the current study. Nevertheless, it is expected that the direction – though not the size – of the BFPLE will generalise across countries.

**Rationale for Hypothesis 1.4: Relation Between Individual Math Ability and Math Self-Concept in Individual Countries**

The 41 countries of the current sample comprised individualist countries, collectivist countries, economically developed countries, and economically developing countries. As noted in the rationale for Hypothesis 1.1, most of the research in this area has been conducted in developed countries and individualist countries and has demonstrated that individual ability has a significantly positive effect on various facets of academic self-concept (e.g., Marsh, 1991; Marsh & Parker, 1984; Marsh et al., in press). There is, however, some evidence to show that a similar relation holds for collectivist countries and developing countries (e.g., Marsh et al., 2000; Zeidner & Schleyer, 1998). Marsh and Hau (2003) demonstrated that all 26 countries in their cross-national study displayed a positive relation between academic ability and academic self-concept. Therefore, it is expected that there will be a significantly positive relation between individual ability and math self-concept in each of the individual countries in the current study.

**Rationale for Hypothesis 1.5: Relation Between School-Average Math Ability and Math Self-Concept in Individual Countries**

The individual countries of the current sample differ culturally and economically, and so different predictions could be formulated regarding the relation between school-average ability and math self-concept (the BFLPE). On the basis of acknowledged differences between individualist countries and collectivist societies (see Chapter 2) it could be argued that individuals from collectivist countries would
not display a BFLPE. Compared to individualist countries, collectivist cultures place more emphasis on the rights and responsibilities of the social group than those of the individual. Students in those countries that value collective principles may not display the negative relation between school-average ability and academic self-concept that characterises the BFLPE, as they may place more value on group achievements than on their own (Marsh et al., 2000). This viewpoint was supported in an experimental study by McFarland and Buehler (1995), in which they demonstrated that the BFLPE was less for participants who hailed from a collectivist background compared to those with an individualist heritage (see Chapter 2).

For developing countries, the relation between school-average ability and academic self-concept may be more complex. In developing countries, where educational standards are lower, there may be no opportunity for the BFLPE to occur, as the academic differences between schools may be minimal. Alternatively, the divide between rich and poor in developing countries may mean that the rich receive higher standards of education more reflective of developed countries, and so the negative relation between school-average ability and academic self-concept, which indicates the BFLPE, may be evident.

These, however, are only intuitive perspectives concerning the extent of the BFLPE in collectivist and developing countries, as there is very little supporting empirical evidence for these viewpoints (but see McFarland & Buehler, 1995; Marsh et al., 2000; Marsh & Hau, 2003; Zeidner & Schleyer, 1998). This is in contrast to considerable empirical research in developed countries and individualist countries that has shown that, when individual ability is controlled, class- and school-average ability have a statistically significant negative effect on academic self-concept (e.g., Craven et al., 2000; Marsh, 2005; Marsh & Craven, 2002; Marsh et al., 2001; Marsh & Rowe, 1996; Marsh et al., in press; Mulkey et al., 2005). Moreover, the little research that has been conducted in collectivist countries and economically developing countries has produced similar results (Marsh et al., 2000; Marsh & Hau, 2003; Zeidner & Schleyer, 1998). Accordingly, a hypothesis stating that there will be a negative relation between school-average math ability and math self-concept has been formulated, based on the strength of actual results rather than intuitive
suppositions, although it is expected that the size of this effect will be different across countries.

**Rationale for Research Question 1.1: Culture as a Moderator of the BFLPE and Research Question 1.2: Economic Development as a Moderator of the BFLPE**

As noted in the rationales for Hypotheses 1.1, 1.2, 1.4, and 1.5, previous research has shown that collectivist countries and economically developing countries displayed a BFLPE (Marsh & Hau, 2003; Marsh et al., 2000; Zeidner & Schleyer, 1998). These studies demonstrated a statistically significant positive relation between individual ability and academic self-concept, and a statistically significant negative relation between class- and school-average ability and academic self-concept – the two components necessary to provide evidence of a BFLPE. These studies demonstrated that whether countries were individualist, collectivist, economically developing, or developed, was irrelevant, as they all displayed the relations that evidence a BFLPE. However, to date no study has investigated the main effect of type of culture or economic development on academic self-concept, nor whether the BFLPE can be moderated by either of these constructs. Although it is anticipated that the main effect of type of culture and economic development and their interactions with school-average ability will not have a substantial association with math self-concept, as previous research cannot provide clear guidelines, research questions have been posed to elucidate these issues.

**Individual Ability and the BFLPE**

**Rationale for Research Question 2.1: Individual Math Ability as a Moderator of the BFLPE**

Evidence surrounding the moderating effect of individual ability (linear component) on the BFLPE has been conflicting. Some studies have shown little or no evidence of a BFLPE for high-ability students in academically selective classes (e.g., Coleman & Fults, 1985). However, Reuman (1989) noted that high-ability students in homogeneous high-ability classes felt the negative effects of the BFLPE more than low-ability students in low-ability classes. Marsh et al. (2000) noted that the interaction effects of ability and school-average ability were generally small and
the directions of the effects varied across models. Other studies have demonstrated that students of all ability levels suffer the negative effects of the BFLPE (e.g., Marsh et al., 1995; Marsh & Hau, 2003; Marsh & Rowe, 1996). As regards the quadratic component of individual ability, although Marsh and Hau (2003) demonstrated that the quadratic ability by school-average ability interaction was significantly positively related to academic self-concept, the size of the effect, given the extremely large sample, was small. Given these conflicting results, no a priori predictions were made; instead, a research question was posed to clarify this matter.

**Individual Student Characteristics and the BFLPE: SES, Individual Differences in Learning, and Perceptions of the Learning Environment**

**Rationale for Hypothesis 3.1: Relation Between Individual Socio-Economic Status (SES) and Math Self-Concept**

Although White (1982) found only a weak correlation between SES and academic performance, more recent studies have shown that SES is positively associated with academic achievement (DeGarmo et al., 1999; Marks et al., 2006; Organisation for Economic Cooperation and Development, 2001a; Pong & Ju, 2000). Additionally, Marsh and Parker (1984) and Bachman and O’Malley (1986) noted that individual SES had a significantly positive effect on academic self-concept. It is expected that the current study will display a similar statistically significant positive relation between SES and math self-concept.

**Rationale for Research Question 3.1: Individual SES as a Moderator of the BFLPE**

To date, no studies have investigated the moderating effect of individual SES on the BFLPE. It may be that high SES students suffer the BFLPE more than low SES students because they feel more pressure to perform well in school and live up to their parents’ expectations. Conversely, high SES students may suffer the BFLPE less than low SES students. The parents of high SES students are more likely to be highly educated. Parents with higher education levels may be more able to provide their children with coping strategies, and so high SES students may be more able to cope with the demands and pressures of high-ability classes and schools than
students from low SES backgrounds. As past research and theory were unable to provide guidance for formulating clear predictions and as SES could affect the BFLPE in opposing ways, a research question was posed to elucidate this issue.

**Rationale for Hypothesis 3.2: Relation Between Self-Regulated Learning Strategies and Math Self-Concept**

Research suggests that more effective learning is promoted by the use of self-regulated learning strategies (such as elaboration, memorisation, and control strategies; Marsh, Hau, et al., 2006), with students who use such strategies performing better academically (Boekaerts, 1997; Boekaerts & Cascallar, 2006; Camahalan, 2006). In the current study it is expected that a similar positive relation will exist between the use of self-regulated learning strategies and math self-concept. Hence, a hypothesis was posed predicting that there would be a statistically significant positive relation between each of the self-regulated learning strategy variables and math self-concept.

**Rationale for Hypothesis 3.3: Relation Between Motivation and Math Self-Concept**

Whereas intrinsic motivation has been associated with positive educational outcomes (e.g., Ginsburg & Bronstein, 1993; Gottfried, 1985, 1990), some studies have indicated that extrinsic motivation is associated with negative educational outcomes (Lepper et al., 2005). However, others (Otis et al., 2005; Ryan & Deci, 2000) have noted that extrinsic motivation can also be associated with positive educational results. Consistent with these latter findings, it is anticipated that both intrinsic and extrinsic motivation will have a significantly positive relation with math self-concept.

**Rationale for Hypothesis 3.4: Relation Between Math Self-Efficacy and Math Self-Concept**

Self-efficacy has been shown to be a significant positive predictor of academic performance (Multon et al., 1991) and to be positively related to math self-concept (Marsh, Hau, et al., 2006). It is expected that a similar positive relation will apply in the current study between self-efficacy and math self-concept. (Note that in the Pisa 2000 database (see Organisation for Economic Cooperation and Development, 2001a, 2001b), the self-efficacy measure was similar to self-concept. However, in the
PISA 2003 database (see Organisation for Economic Cooperation and Development, 2005a, 2005b) used in the current study, the self-efficacy measure is quite different to self-concept, being based more on absolute standards. See the example item in Table 4. 3.

**Rationale for Hypothesis 3.5: Relation Between Math Anxiety and Math Self-Concept**

Studies have indicated that there is a negative relation between anxiety and academic performance (Chapell et al., 2005; Martin, 2003; Zeidner & Schleyer, 1998) and that anxiety is negatively related to math ability perception, a concept similar to math self-concept (Meece et al., 1990). Thus it is anticipated that a similar negative relation between math anxiety and math self-concept will exist in the current study.

**Rationale for Hypothesis 3.6: Relation Between Preference for Learning Environment and Math Self-Concept**

Studies comparing academic achievement in cooperative and competitive learning environments have produced conflicting results (Sherman, 1988; Slavin, 1983). However, when preferences for a particular learning environment have been considered, both types of learning situations (cooperative and competitive) were associated with better performance (Organisation for Economic Cooperation and Development, 2001a). A similar result is expected in the current study with a preference for cooperative and competitive environments both being positively associated with math self-concept.


As none of these constructs has been previously considered in relation to the BFLPE, the exact nature of the effects they may produce is open to speculation (see Chapter 2 for a discussion of these speculations for each construct considered). For example, anxiety has been associated with poorer performance (e.g., Ironsmith et al., 2003) and so highly anxious students may not perform as well academically as they feel they should, causing them to feel more negative about their abilities. Attending a
high-ability school may cause anxious students even more anxiety, thus contributing to the BFLPE. Supporting evidence for this viewpoint comes from research examining the relation between math ability perception (measured similarly to math self-concept) and math anxiety that has displayed a negative relation between the two constructs (Meece et al., 1990). However, the ways in which anxiety and the other constructs considered (self-regulated learning strategies, motivation, self-efficacy, and preference for learning environment) might affect the BFLPE can be conjecture only. As there is no empirical evidence or any theory on which to base hypotheses for these constructs, research questions have been posed to illuminate whether they alleviate or exacerbate the BFLPE.

**Rationale for Hypothesis 3.7: Relation Between Individual Perceptions of the Learning Environment and Math Self-Concept**

Positive perceptions of the learning environment have been related to better educational and mental health outcomes (Crosnoe et al., 2004; Ozer, 2005). It is anticipated that a similar relation will be found in the current study, such that individual perceptions of the learning environment will be positively related to math self-concept.

**Rationale for Research Question 3.7: Individual Perceptions of the Learning Environment as Moderators of the BFLPE**

Individual students’ perceptions of their learning environment have not been explored in relation to the BFLPE. It may be that students who hold positive perceptions of their school will not be affected by the BFLPE. For example, high-ability students in high-ability schools may feel more of a sense of belonging to their school because they do not feel different to other students academically. They may suffer the BFLPE less than others who do not feel this sense of belonging simply because they feel a connection to their school. However, no research exists on which to base a hypothesis and so a research question was devised to ascertain whether individual perceptions of the learning environment could moderate the BFLPE.
Section Summary

This section has provided a brief rationale for each of the hypotheses and research questions to be examined in the current study. These rationales were based on pertinent literature reviewed in Chapters 2 and 3.

Methodology

Participants

More than a quarter of a million (N=276,165) fifteen-year-old students from 41 countries participated in the PISA study conducted by the OECD in the year 2003 (see Organisation for Economic Cooperation and Development, 2005a, 2005b). These students completed paper-and-pencil tests to assess their knowledge and skills in reading, math, science, and problem solving. Rather than assessing curricula mastery, these tests were designed to ascertain how well these students, nearing the end of their formal education, had mastered the skills necessary for life after school. The tests included both multiple choice items, short answer questions, and extended response items. Different combinations of the assessment items were administered to students. To accomplish this, 13 booklets of test items were created, each with a different set, or cluster, of items. Thus, Booklet 1 contained 3 sets of math items and one set of reading items, while Booklet 10 contained one set of math items, one of problem solving, and two sets of reading items. All booklets contained at least one set of math items. Booklets were randomly allocated to students (one booklet per student), and testing lasted for a total of two hours (Organisation for Economic Cooperation and Development, 2005b).

In addition, each student completed a background questionnaire that assessed a variety of areas including perceptions of learning environment, SES, individual student approaches to learning, and the central focus of this study, math self-concept. This questionnaire took approximately 30 minutes to complete. Countries were also given the option for their students to complete questionnaires regarding their educational careers and their familiarity with computers.
The PISA tests are administered every three years, each year having a different academic focus; for the 2003 administration it was math. However, not all students completed the math self-concept items that were central to the current study and so these students were deleted from further analyses. Additionally, to be comparable with the earlier Marsh and Hau (2003) study, schools with 10 participating students or less (917 schools, 8.9% of total schools) were also deleted from further analyses, as these schools were considered to be too small to be included in multilevel analyses. This resulted in a sample of 265,180 students, in 10,221 schools, across 41 countries for the purposes of the present investigation.

Measures

Math Ability

As noted in the PISA documentation (Organisation for Economic Cooperation and Development, 2005a), measures used in the psychological or educational realm often encompass measurement error for a variety of reasons. For instance, in educational research results could be affected by how the students feel both mentally and physically on the day of the testing, or by the physical conditions under which the students are tested (see Organisation for Economic Cooperation and Development, 2005a, pp. 72 – 80). Additionally, the purpose of PISA was “To assess the knowledge or skills of a population” (Organisation for Economic Cooperation and Development, 2005a, p. 72) and so the more important goal was to reduce measurement error at the target population level rather than at the individual level. For these reasons the PISA database does not contain a single math ability measure. Rather, to prevent biased population estimates being obtained, the PISA database used five plausible values to estimate a student’s academic ability. The PISA documentation quotes Wu and Adams’ (2002) description of plausible values as a “representation of the range of abilities a student might reasonably have...Instead of directly estimating a student’s ability \( \theta \), a probability distribution for a student’s \( \theta \) is estimated” (Organisation for Economic Cooperation and Development, 2005a, p. 75). The PISA documentation warns against averaging these plausible values or using a single plausible value to estimate an individual’s ability. Rather, researchers are advised that statistics should be based on each plausible value separately and then
those statistics should be averaged. In the current study, each analysis was conducted with all five plausible values separately and all resulting parameters were averaged (see Organisation for Economic Cooperation and Development, 2005a). Standard errors were calculated according to set criteria for dealing with plausible values for multilevel analyses (see Organisation for Economic Cooperation and Development, 2005a and Raudenbush, Bryk, & Congdon, 2005). These calculations are described in detail in Appendix A.

**Cultural Orientation: Individualism / Collectivist Index**

Countries were scaled along an individualist/collectivist dimension using the Individualism Index scale, developed by Hofstede and Hofstede (2005). Scores on this scale ranged from 14 to 91, with a higher score meaning that the country was higher in individualism, and a lower score meaning that the country was more collectivist in culture. The scale score assigned to each country is presented in Table 4.1. The mean score on this index was 57.05 ($SD = 23.55$). Table 4.1 indicates that scores were spread reasonably consistently across the entire range (20 countries scored above the mean and 16 below). Thus individualist and collectivist countries were both substantially represented in the current sample. Five countries in the current sample were not represented in the Hofstede and Hofstede scale and these are noted in Table 4.1. These five countries were not included in tests of the moderating analyses of cultural orientation on the BFLPE. Scale scores were subsequently standardised for multilevel analyses.

**Developing / Developed Countries**

The World Bank (2007) classifies countries into four economic groups according to their gross national income per capita: low income, lower middle income, upper middle income, and high income. Table 4.2 lists the countries included in these analyses and indicates the World Bank classification. According to the World Bank, low and middle-income economies can also be referred to as developing economies. For purposes of the present investigation countries were coded as either developing (coded 0) or developed (coded 1), which resulted in 27 countries being classified as developed and 14 as developing.
Table 4.1. *Individualist/Collectivist Index Score by Country* (based on Hofstede & Hofstede, 2005)

<table>
<thead>
<tr>
<th>Country</th>
<th>Individualism score</th>
<th>Country</th>
<th>Individualism score</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>91</td>
<td>Slovak republic</td>
<td>52</td>
</tr>
<tr>
<td>Australia</td>
<td>90</td>
<td>Spain</td>
<td>51</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>89</td>
<td>Japan</td>
<td>46</td>
</tr>
<tr>
<td>Canada</td>
<td>80</td>
<td>Russian Federation</td>
<td>39</td>
</tr>
<tr>
<td>Hungary</td>
<td>80</td>
<td>Brazil</td>
<td>38</td>
</tr>
<tr>
<td>Netherlands</td>
<td>80</td>
<td>Turkey</td>
<td>37</td>
</tr>
<tr>
<td>New Zealand</td>
<td>79</td>
<td>Uruguay</td>
<td>36</td>
</tr>
<tr>
<td>Italy</td>
<td>76</td>
<td>Greece</td>
<td>35</td>
</tr>
<tr>
<td>Belgium</td>
<td>75</td>
<td>Mexico</td>
<td>30</td>
</tr>
<tr>
<td>Denmark</td>
<td>74</td>
<td>Portugal</td>
<td>27</td>
</tr>
<tr>
<td>France</td>
<td>71</td>
<td>Hong Kong</td>
<td>25</td>
</tr>
<tr>
<td>Sweden</td>
<td>71</td>
<td>Macao (China)</td>
<td>20</td>
</tr>
<tr>
<td>Ireland</td>
<td>70</td>
<td>Thailand</td>
<td>20</td>
</tr>
<tr>
<td>Norway</td>
<td>69</td>
<td>Korea</td>
<td>18</td>
</tr>
<tr>
<td>Germany</td>
<td>67</td>
<td>Indonesia</td>
<td>14</td>
</tr>
<tr>
<td>Switzerland</td>
<td>67</td>
<td>Iceland</td>
<td>No score</td>
</tr>
<tr>
<td>Finland</td>
<td>63</td>
<td>Latvia</td>
<td>No score</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>60</td>
<td>Liechtenstein</td>
<td>No score</td>
</tr>
<tr>
<td>Poland</td>
<td>60</td>
<td>Tunisia</td>
<td>No score</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>58</td>
<td>Serbia and Montenegro</td>
<td>No score</td>
</tr>
<tr>
<td>Austria</td>
<td>55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Countries are presented in order of Individualism / Collectivist Index score, from the highest to the lowest score. A high score indicates a high level of individualism. “No score” indicates that there was no Individualism / Collectivist score available for that country. Mean score = 57.05.
### Table 4.2. Economic Classification Based on World Bank (2007) Categories

<table>
<thead>
<tr>
<th>Country</th>
<th>Economic Classification</th>
<th>Country</th>
<th>Economic Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>High (1)</td>
<td>Luxembourg</td>
<td>High (1)</td>
</tr>
<tr>
<td>Austria</td>
<td>High (1)</td>
<td>Macao (China)</td>
<td>High (1)</td>
</tr>
<tr>
<td>Belgium</td>
<td>High (1)</td>
<td>Mexico</td>
<td>Upper middle (0)</td>
</tr>
<tr>
<td>Brazil</td>
<td>Lower Middle (0)</td>
<td>Netherlands</td>
<td>High (1)</td>
</tr>
<tr>
<td>Canada</td>
<td>High (1)</td>
<td>New Zealand</td>
<td>High (1)</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>Upper Middle (0)</td>
<td>Norway</td>
<td>High (1)</td>
</tr>
<tr>
<td>Denmark</td>
<td>High (1)</td>
<td>Poland</td>
<td>Upper Middle (0)</td>
</tr>
<tr>
<td>Finland</td>
<td>High (1)</td>
<td>Portugal</td>
<td>High (1)</td>
</tr>
<tr>
<td>France</td>
<td>High (1)</td>
<td>Russian Federation</td>
<td>Upper Middle (0)</td>
</tr>
<tr>
<td>Germany</td>
<td>High (1)</td>
<td>Slovak Republic</td>
<td>Upper middle (0)</td>
</tr>
<tr>
<td>Greece</td>
<td>High (1)</td>
<td>Spain</td>
<td>High (1)</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>High (1)</td>
<td>Sweden</td>
<td>High (1)</td>
</tr>
<tr>
<td>Hungary</td>
<td>Upper Middle (0)</td>
<td>Switzerland</td>
<td>High (1)</td>
</tr>
<tr>
<td>Iceland</td>
<td>High (1)</td>
<td>Thailand</td>
<td>Lower Middle (0)</td>
</tr>
<tr>
<td>Indonesia</td>
<td>Lower Middle (0)</td>
<td>Tunisia</td>
<td>Lower Middle (0)</td>
</tr>
<tr>
<td>Ireland</td>
<td>High (1)</td>
<td>Turkey</td>
<td>Upper Middle (0)</td>
</tr>
<tr>
<td>Italy</td>
<td>High (1)</td>
<td>United Kingdom</td>
<td>High (1)</td>
</tr>
<tr>
<td>Japan</td>
<td>High (1)</td>
<td>United States</td>
<td>High (1)</td>
</tr>
<tr>
<td>Korea</td>
<td>High (1)</td>
<td>Uruguay</td>
<td>Upper Middle (0)</td>
</tr>
<tr>
<td>Latvia</td>
<td>Upper middle (0)</td>
<td>Serbia and Montenegro</td>
<td>Lower Middle (0)</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>High (1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Countries with an upper or lower middle economic classification were denoted as developing (coded as 0). Countries with a high economic classification were denoted as developed (coded as 1).

### PISA 2003 Indices

Using theory and previous research, responses to questions in the background questionnaire were combined to create scales for various constructs. Two types of scales, or indices as they are referred to in the PISA documentation, were used in the PISA database: simple and scale indices. Simple indices were created by arithmetically transforming or recoding items and meaningful variables were calculated using item responses. In the current study, two of the SES variables (highest occupational status of parents, and highest educational level of parents) were simple indices. All other indices were scale indices. For these scale indices, item response theory was utilised to construct estimates of latent traits. Structural
Equation Modelling (SEM) and Confirmatory Factor Analysis (CFA) were used to validate the scale indices by confirming “theoretically expected dimensions and, if necessary, to re-specify the dimensional structure” (Organisation for Economic Cooperation and Development, 2005b, p. 282). Where necessary, items were inverted for scaling. For all indices higher values signified higher levels of the index under discussion. For example, a high score on the index of cooperative learning indicated the student’s preference for a more cooperative learning environment. All indices were standardised to have a mean of 0 and a standard deviation of 1 across the entire sample. Those indices used in the present investigation are described further below (for full details of indices used in the PISA database, see Organisation for Economic Cooperation and Development, 2005b, pp. 271-320).

**Math self-concept.** Math self-concept was measured in the PISA 2003 database (see Organisation for Economic Cooperation and Development, 2005a, 2005b) using five items, for example, “I get good marks in mathematics” and “I learn quickly in maths”. These items were scored on a 4-point Likert scale ranging from 1 (*strongly agree*) to 4 (*strongly disagree*). Four items were inverted for scoring so that a high score reflected a higher math self-concept. The reliability of this scale in the current sample was high, with a Cronbach’s alpha of .88. This variable was standardised (*M* = 0, *SD* = 1) across the entire sample and scores ranged from -2.24 to 2.48.

**Socio-economic status.** The PISA 2003 database (see Organisation for Economic Cooperation and Development, 2005a, 2005b) offered various indices to measure SES. These included indices based on answers students gave to questions about parental occupation, parental education, and possessions within the home.

The parental occupation measure was based on whichever occupational status was higher, the mother’s or father’s. It was assessed using open-ended questions, which were first coded according to the International Standard Classification of Occupations (ISCO; International Labour Office, 1990) and then recoded into the International Socio-Economic Index of Occupational Status (ISCO; Ganzeboom, De Graaf, & Treiman, 1992; refer Organisation for Economic Cooperation and Development, 2005b). Higher scores represent higher levels of parental income. When standardised across the entire sample to have a mean of 0 and a standard deviation of 1, this index ranged from -1.82 to 2.49.
The parental education measure was based on whichever educational level was higher, the mother’s or father’s. It was classified according to the International Standard Classification of Education (ISCED; Organisation for Economic Cooperation and Development, 1999) so that educational levels were comparable across countries. Six categories were used ranging from 0 (None) to 6 (Theoretically orientated tertiary and postgraduate). Hence, higher scores represent higher levels of parental education. This index was standardised ($M = 0$, $SD = 1$) across the entire sample for subsequent analyses. It ranged from -2.36 to 1.16.

The home educational resources scale consisted of five items such as, “In your home do you have books to help with your schoolwork?” Three items comprised the cultural possessions scale. Items included, “In your home do you have books of poetry?” Items on both these scales were answered by either “yes” or “no”. All items were inverted for scaling, meaning that higher scores reflected higher SES levels. Reliabilities for these scales in the current study ranged from .58 for the home educational resources scale to .67 for the cultural possessions scale. All scales were standardised to have a mean of 0 and a standard deviation of 1. The range of standardised scores in each scale was: home educational resources = -.388 to .74, and cultural possessions = -1.18 to 1.44.

The PISA 2003 database (see Organisation for Economic Cooperation and Development, 2005a, 2005b) also provided a combined SES variable that was an index of economic, social, and cultural status. It was derived from three of the PISA indices: highest parental education, highest parental occupation, and the number of possessions in the home. The rationale for using these three measures was that SES is typically based on education, occupational status and income. However, as there was no direct measure of income in the PISA database, household possessions were considered to be “an approximate measure of family wealth” (Organisation for Economic Cooperation and Development, 2005b, p. 316). This index was standardised across the entire sample to have a mean of 0 and a standard deviation of 1, and ranged from -4.14 to 2.92.

**Individual differences in learning.** Individual differences in learning were measured using nine of the PISA indices that assessed four broad theoretical dimensions: self-regulated learning strategies, motivation to learn, self-related
cognitions, and learning preferences. Details of the factors used, including number of items, an example item, reliability, the response scales, range of scores and the meaning of a high score are presented in Table 4.3. All indices were standardised across the entire sample ($M = 0$, $SD = 1$).

**Individual perceptions of the learning environment.** The PISA 2003 database (see Organisation for Economic Cooperation and Development, 2005a, 2005b) includes “three student-level indices”, which assessed the learning environment as perceived by the individual student (see Organisation for Economic Cooperation and Development 2005b, p. 285). These indices were student teacher relations, a sense of belonging to the school, and attitudes to school. Details, including number of items, an example item, reliability, the response scales, range of scores, and the meaning of a high score are presented in Table 4.4. All indices were standardised across the entire sample to have a mean of 0 and a standard deviation of 1.
Table 4.3. **Summary of Key Features of the Individual Differences in Learning Indices**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Index</th>
<th>No. of Items</th>
<th>Cronbach’s α</th>
<th>Example Item</th>
<th>Response Scale</th>
<th>Range of Standardised Scores</th>
<th>Meaning of a High Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Strategies</td>
<td>Control Strategies</td>
<td>5</td>
<td>.74</td>
<td>“When I study for a mathematics test, I try to work out what are the most important parts to learn”</td>
<td>4-point Likert scale ranging from 1 <em>(strongly agree)</em> to 4 <em>(strongly disagree)</em>. All items inverted for scaling.</td>
<td>-3.57 – 2.64</td>
<td>A high score indicates a preference for this learning strategy.</td>
</tr>
<tr>
<td>Memorisation</td>
<td>4</td>
<td>.59</td>
<td>“When I study for mathematics, I try to learn the answers to problems off by heart”</td>
<td>4-point Likert scale ranging from 1 <em>(strongly agree)</em> to 4 <em>(strongly disagree)</em>. All items inverted for scaling.</td>
<td>-3.62 – 3.22</td>
<td>A high score indicates a preference for this learning strategy.</td>
<td></td>
</tr>
<tr>
<td>Elaboration</td>
<td>5</td>
<td>.76</td>
<td>“When learning mathematics, I try to relate the work to things I have learnt in other subjects”</td>
<td>4-point Likert scale ranging from 1 <em>(strongly agree)</em> to 4 <em>(strongly disagree)</em>. All items inverted for scaling.</td>
<td>-3.41 – 3.10</td>
<td>A high score indicates a preference for this learning strategy.</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3. Summary of Key Features of the Individual Differences in Learning Indices (contd.)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Index</th>
<th>No. of Items</th>
<th>Cronbach’s α</th>
<th>Example Item</th>
<th>Response Scale</th>
<th>Range of Standardised Scores</th>
<th>Meaning of a High Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Extrinsic</td>
<td>4</td>
<td>.88</td>
<td>“Making an effort in mathematics is worth it because it will help me in the work I want to do later on”</td>
<td>4-point Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree). All items inverted for scaling.</td>
<td>-2.54 – 1.67</td>
<td>A high score indicates higher levels of instrumental motivation.</td>
</tr>
<tr>
<td></td>
<td>Intrinsic</td>
<td>4</td>
<td>.89</td>
<td>“I am interested in the things I learn in mathematics”</td>
<td>4-point Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree). All items inverted for scaling.</td>
<td>-1.94 – 2.30</td>
<td>A high score indicates higher levels of interest and enjoyment in math.</td>
</tr>
<tr>
<td>Self-related</td>
<td>Math Self-Efficacy&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>8</td>
<td>.82</td>
<td>“How confident do you feel about …. solving an equation like 2(x+3) = (x+3)(x-3)?”</td>
<td>4-point Likert scale ranging from 1 (very confident) to 4 (not at all confident). All items inverted for scaling.</td>
<td>-4.00 – 2.64</td>
<td>A high score indicates confidence with mathematical tasks.</td>
</tr>
<tr>
<td></td>
<td>Math Anxiety</td>
<td>5</td>
<td>.81</td>
<td>“I get very nervous doing mathematics problems”</td>
<td>4-point Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree). All items inverted for scaling.</td>
<td>-2.70 – 2.72</td>
<td>A high score indicates a higher level of math anxiety.</td>
</tr>
</tbody>
</table>
Table 4.3. *Summary of Key Features of the Individual Differences in Learning Indices (contd.)*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Index</th>
<th>No. of Items</th>
<th>Cronbach’s α</th>
<th>Example Item</th>
<th>Response Scale</th>
<th>Range of Standardised Scores</th>
<th>Meaning of a High Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Preference</td>
<td>Co-operative Learning</td>
<td>5</td>
<td>.78</td>
<td>“In mathematics I enjoy working with other students in groups”</td>
<td>4-point Likert scale ranging from 1 <em>(strongly agree)</em> to 4 <em>(strongly disagree)</em>. All items inverted for scaling.</td>
<td>-3.29 – 2.73</td>
<td>A high score indicates a preference for a cooperative learning environment.</td>
</tr>
<tr>
<td></td>
<td>Competitive Learning</td>
<td>5</td>
<td>.84</td>
<td>“I would like to be the best in my class in mathematics”</td>
<td>4-point Likert scale ranging from 1 <em>(strongly agree)</em> to 4 <em>(strongly disagree)</em>. All items inverted for scaling.</td>
<td>-3.04 – 2.39</td>
<td>A high score indicates a preference for a competitive learning environment.</td>
</tr>
</tbody>
</table>

*In the Pisa (2000) database, the self-efficacy measure was akin to self-concept. However, in the PISA (2003) database used in the current study, the self-efficacy measure is quite different to self-concept, being based more on absolute standards as shown in the example item.*
<table>
<thead>
<tr>
<th>Index</th>
<th>No. of Items</th>
<th>Cronbach’s α</th>
<th>Example Item</th>
<th>Response Scale</th>
<th>Range of Standardised Scores</th>
<th>Meaning of a High Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Teacher Relations</td>
<td>5</td>
<td>.79</td>
<td>“Students get along well with most teachers”</td>
<td>4-point Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree). All items inverted for scaling.</td>
<td>-3.15 – 2.72</td>
<td>A high score indicates more positive perceptions of student teacher relations.</td>
</tr>
<tr>
<td>Sense of Belonging</td>
<td>6</td>
<td>.78</td>
<td>“I make friends easily”</td>
<td>4-point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). Three items inverted for scaling.</td>
<td>-3.40 – 2.27</td>
<td>A high score indicates that the student feels more connected to the school.</td>
</tr>
<tr>
<td>Attitudes to School</td>
<td>4</td>
<td>.62</td>
<td>“School helped give me confidence to make decisions”</td>
<td>4-point Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree). Two items inverted for scaling.</td>
<td>-3.18 – 2.36</td>
<td>A high score indicates a more positive attitude to school.</td>
</tr>
</tbody>
</table>
Statistical Analysis – Multilevel Modelling

“A multilevel model is a statistical model applied to data collected at more than one level in order to elucidate relationships at more than one level” (Luke, 2004, pp. 7-8).

So Why Use Multilevel Modelling?

Much of what is studied in psychological and educational research possesses a multilevel structure. People are social animals. Social interaction is not only enjoyable, but it is also necessary for optimal human development (Goldfarb, 1955). Individuals do not live in a vacuum. They belong to families, work with others, belong to social and recreational groups, and are part of larger communities. Inevitably then, individuals are studied at various levels: within their family units, within their socio-economic status, and within their cultural background. Students are also studied at various levels: within their classes and within their schools. Even experimental studies can be thought of as having a multilevel structure: Participants are often arranged in groups and individuals within these groups are administered equivalent treatments. Data that possess this type of structure have been termed nested, multilevel, or hierarchical data, and these terms are used interchangeably. Very often researchers ignore multilevel data when performing statistical analyses, or at the very best incorporate separate variables to represent the different levels. This procedure is inefficient because of the sheer number of extra variables needed to estimate different levels (e.g., if there were 50 schools, 49 dummy variables would be necessary), and inadequate because variation between the groups cannot be observed (Rasbash et al., 2002).

Members of groups tend to share common characteristics. For example, family members tend to be more alike physically and psychologically than members of another family chosen at random. Siblings often have similar hair and eye colour, psychological problems such as depression can run in families, and parents can pass genetic diseases on to their offspring. Participants in the current thesis were members of groups. They were students who attended the same class, the same school, or lived in the same country. So, for instance, students in a rural school in Australia may share more similarities with each other than they do with students from a city school in Australia. Additionally, Australian students, whether they attend rural or city
schools, may share more similarities with each other than they do with, say, Russian students. It becomes clear that participants within groups are not independent of each other, as members of one group tend to be more similar to each other than they are to members of another group. Thus, one of the fundamental assumptions of single-level statistical techniques – that observations are independent – is violated if a multilevel structure is ignored. Traditional single level techniques, such as multiple regression, are inadequate for analysing multilevel data for theoretical and statistical reasons.

**Theoretical reasons.** When researchers use single level techniques to analyse multilevel data, their interpretations can be misleading, especially if they assume that relations discovered at one level apply in a similar way to some other level (Luke, 2004). An example of this occurs when studies observe relations in groups and researchers assume that these relations hold for individuals. For instance, a researcher may want to ascertain the effect of teacher enthusiasm on student achievement. If the data are aggregated to a class level to determine whether the classes differ on their mean achievement scores depending on teacher enthusiasm, then interpretation is limited to the class level. However, often the class results are applied to the individual student level. This misinterpretation of the results is termed an ecological fallacy (Tabachnick & Fidell, 2007). Conversely, data collected at the individual level cannot be said to apply to the groups to which those individuals belong. Were health and socio-economic data to be collected at an individual level across the country, one could not conclude that the effect on health of low socio-economic status for individuals living in one part of the country would be the same as that for individuals living in a completely different region (Rowe, 2005). To do so would be to commit an atomistic fallacy, making inferences about a group from individual-level information (Luke, 2004). Multilevel modelling can resolve these difficulties by modelling the effects at all levels simultaneously and by comparing the variance explained by all the different levels (Rowe, 2005).

**Statistical reasons.** Ignoring a multilevel structure can also result in technical difficulties that can broadly be categorised as aggregation bias, misestimated standard errors, and undetected heterogeneity of regression (Raudenbush & Bryk, 2002). An aggregation bias occurs when a variable can be measured at different levels, but its meaning and effects are different depending on the level. For example,
in organisational research a researcher may measure morale at the individual level, but also at the company level. The average morale of the company may have an effect on the productivity of an individual worker over and above the effect of the individual worker’s morale. At the individual level, morale provides an indication of the resources available to the individual. At the company level, morale is a proxy measure of the company’s resources. Multilevel modelling can overcome this difficulty by separating effects into higher (e.g., company) and lower (e.g., individual worker) order components (Raudenbush & Bryk, 2002; Rowe, 2005).

Single level multiple regression assumes that the data are independent. However, because members of one group are more similar to each other than they are to members of another group, this assumption of independence is often violated. For example, in educational research, students within a particular school will be more alike (e.g., from similar socio-economic backgrounds), on average, than they are to students from another school, and thus their responses will be correlated (Rowe, 2005). Additionally, students in a class are likely to influence each other, but such influence may not permeate the school as a whole. This lack of independence typically results in an underestimation of the standard errors for the higher level variables leading to the likelihood of finding a significant effect where none exists. Multilevel models resolve this problem by including a unique random effect for every group (e.g., class or school) in the statistical model (Raudenbush & Bryk, 2002; Rowe, 2005).

If undetected, heterogeneity of regression, when relations between individual characteristics and outcomes vary across groups or organisations, is another problem faced when data have a hierarchical structure. Rowe (2005) cites school-based studies that demonstrate that the way in which schools organise their administration and social features are important factors that affect teachers and students alike (e.g., Hill & Rowe, 1996; Lee, Dedrick, & Smith, 1991). Multilevel modelling accounts for differences such as these by computing separate regression lines for each school, thus modelling any variation that exists between schools (Rowe, 2005).
**Fitting Multilevel Models**

Over the past 20 years, multilevel models, also known as multilevel linear models, random-effects models, and hierarchical linear models, have grown in popularity, being used in such disciplines as economics, education, epidemiology, and biology. Unfortunately, psychologists have been slow to recognise the advantages that multilevel models can bring to their discipline (Rowe, 2005). Educational psychologists, however, have acknowledged the importance of accounting for multilevel structures. Marsh and Hau’s (2003) analysis of the BFLPE across 26 countries, described in Chapter 2, is one example of multilevel modelling being employed to analyse data containing a multilevel structure within educational psychology. As all three studies in the present investigation contained a multilevel structure, multilevel modelling was used for analyses. In Study 1, students were nested within schools within countries; in Studies 2 and 3, students were nested within classes within schools.

The multilevel models used in the present investigation are multilevel regression models that in essence are a multilevel version of the well-known regression model. Just as in multiple regression, a multilevel regression model can be used to analyse a variety of research problems and can accommodate both continuous and categorical variables (Hox, 2002). However, the multilevel models used in the present investigation were idiosyncratic to each study, so no one model can be described herein as representative of the models used. Hence, to illustrate how multilevel modelling incorporates variation between groups in the present investigation, an example, adapted from Rasbash et al. (2002) and based on data from Study 1 of this thesis, is herewith presented. The example begins with the formulae underlying multilevel modelling and continues by describing the steps used to fit a multilevel model. The sample used for this illustration consists of 12,383 Australian students. The outcome variable is a measure of math self-concept, and for simplicity only one predictor variable, a measure of individual math ability, is included. The focus of interest is in predicting values in the outcome variable given certain values of the predictor variable.
So, for example, a simple research question could be “Does a student’s level of math ability have an effect on his/her math self-concept?” The ordinary single-level regression equation for one predictor variable would be:

\[ y_i = a + bx_i + e_i \] (4.1)

where

- \( y \) = predicted value of the outcome variable
- \( a \) = the intercept, where the regression line crosses the \( y \) axis
- \( x \) = predictor variable
- \( b \) = the slope of the regression line, or regression coefficient
- \( i \) = a subscript for individuals that takes the value 1 to the number of students in the sample, in this case, 1 to 12383
- \( e \) = the error term for level 1, the extent to which the \( i \)-th pupil’s actual self-concept score test deviates from the predicted score.

The expression \( a + bx \) is referred to in multilevel modelling as the fixed part of the model. If ordinary regression analysis were to be used to analyse these data, the regression line predicting math self-concept from individual math ability would be similar to the regression line depicted in Figure 4.1.

![Figure 4.1. Predicted Regression Line for a Single-Level Model](image)

Of course, this model does not account for the fact that the students in this sample belong to more than one school, or that the schools may have different intercepts. The 12,383 students in the sample were in 318 schools, so for multilevel
modelling it is necessary to add a regression line for each school. This equation can be written concisely as:

\[
\hat{y}_{ij} = a_j + bx_{ij}
\]  

(4.2)

where \( j \) = a subscript for classes that takes the value 1 to the total number of schools sampled, in this case, 1 to 318. Hence the full model can be expressed as:

\[
y_{ij} = a_j + bX_{ij} + e_{ij}
\]  

(4.3)

However, level 2 variables (schools in this case) are treated by multilevel analysis as a random sample from the population, so an error term for level 2 is necessary. Hence, the equation (4.2) can be re-expressed as

\[
a_j = a + u_j \\
\hat{y}_{ij} = a + bx_{ij} + u_j
\]  

(4.4)

where \( a \) = a constant

\( u_j \) = a level 2 error term, the extent to which the \( j \)-th school’s intercept deviates from the overall intercept value.

The model for actual scores for the two levels can now be written as:

\[
y_{ij} = a + bx_{ij} + u_j + e_{ij}
\]  

(4.5)

By attaching the \( j \) subscript, it is assumed that each school has a different intercept coefficient and a different slope coefficient. There is also now an extra error component in the model - \( u_j \). It is the addition of the \( j \) subscripts, and by extension the extra error term \( u_j \), that sets the multilevel regression equation apart from the ordinary regression equation (Hox, 2002). Hence, in addition to the fixed part of the model – \( a + bx \) – there is also an error component, expressed as \( u_j + e_{ij} \). In multilevel modelling, this error component is called the random part of the model (Rashbash et al., 2002).

The first step in fitting a multilevel model is to define the outcome variable and to state how many levels are involved in the model. In the current example, two
levels were specified. The higher level, level 2, was the school level, and the lower level, level 1, was the individual student level. (In the three studies of the present investigation there were typically two or three levels in each model, and these are outlined within the description of each study). The two-level model, regressing individual math ability on math self-concept, is conveyed by the multilevel modelling statistical software package MLwiN (Rasbash et al., 2002) as:

\[
\begin{align*}
\text{zscmat}_{ij} & \sim N(X\beta, \Omega) \\
\text{zscmat}_{ij} &= \beta_0 + \nu_{ij} + \epsilon_{ij} \\
\beta_{ij} &= \beta_0 + \nu_{ij} + \epsilon_{ij} \\
\begin{bmatrix} \nu_{ij} \\ \epsilon_{ij} \end{bmatrix} & \sim N(0, \Omega) : \Omega = \begin{bmatrix} \sigma^2_{\nu} \\ \sigma^2_{\epsilon} \end{bmatrix}
\end{align*}
\]

In the first line, zscmat is the outcome variable, math self-concept, and N(XB, Ω) is standard notation for the default distributional assumption of normality for the fixed part of the model (XB) and the random component (Ω). In the second line, the equation is defined. β₁ is the regression coefficient associated with the predictor variable zpv1math, a fixed component denoting individual math ability. β₀ is the intercept coefficient associated with the variable CONS. This variable, necessary for every equation, is comprised of a constant vector of ones and is always allowed to vary. It is called CONS by convention. In line three β₀ is further defined. It has three components: a fixed component and two random components associated with the two levels defined in the model. β₀, the coefficient for the “total group” intercept, is the fixed component of the intercept. The random part of the intercept is divided into two parts, sometimes called variance components or random effects: u₀ represents the variability between schools (i.e., how much each school intercept varies from the total group intercept – the school level random effects). Note that this only has a j subscript as it only varies according to school. e₀ represents the variability between individual students (i.e., how much the intercept for the individual student varies from the total group – the individual student random effects). This latter term has both an i and a j subscript as it varies according to both school and student. The coefficients for u₀ and e₀ are presented in lines four and
five when the model has successfully converged. Running this model results in the following output:

\[
\begin{align*}
z_{\text{scmat}, j} & \sim N(\gamma_{0}, \Omega_{0}) \\
z_{\text{scmat}, j} & = \beta_{00} \text{CONS} + 0.447(0.009)z_{\text{pv1math}, j} \\
\beta_{0j} & = -0.060(0.012) + \mu_{0j} + \theta_{0j} \\
\begin{bmatrix} \mu_{0j} \\ \theta_{0j} \end{bmatrix} & \sim N(0, \Omega_{\theta}) : \Omega_{\theta} = \begin{bmatrix} 0.025(0.003) \\ 0.713(0.009) \end{bmatrix} \\
-2*\text{loglikelihood(TOLLS Deviance)} & = 31229.060(12383 \text{ of 12383 cases in use})
\end{align*}
\]

MLwiN uses Maximum Likelihood iterative procedures to produce parameter estimates and standard errors. The parameter estimate (or beta coefficient) for individual math ability (z_{\text{pv1math}}), relating individual math ability to math self-concept, is .447. Its standard error is .009. Using the parameter estimate and the standard error, a Wald test (Wald, 1943) can be used to determine the significance of the parameter estimate. These significance tests are conducted in the form of \( Z = \frac{\text{parameter estimate}}{\text{standard error of the estimate}} \), where \( Z \) is referred to the standard normal distribution (Hox, 2002). If the result of this calculation is greater than 1.96 (or 2 for ease of calculation), then the parameter estimate is statistically significant at the .05 level. In the current example, the parameter estimate for individual math ability (z_{\text{pv1math}}) was highly significant (\( z = \frac{447}{.009} > 1.96 \)).

The total group intercept \( \beta_{0} \) was also significant (\( z = \frac{-0.060}{.012} > 1.96 \)), as were the random effects for the intercept at both level 1 (variance of \( e_{0ij} = \frac{.713}{.009} > 1.96 \)) and level 2 (variance of \( u_{0j} = \frac{.025}{.003} > 1.96 \)). Figure 4.2 displays a graph based on the multilevel model just described. The school slopes are identical as this model only allowed the intercepts to vary. As such, it is called a random intercepts model.
However, schools may also vary in their slopes. This implies that the coefficient of individual math ability will vary from school to school. Running a model in which individual math ability is allowed to vary across schools results in the following output:

$$z_{\text{scmat},y} \sim N(X\beta, \Omega)$$

$$z_{\text{scmat},y} = \beta_{0y} \cdot \text{CONS} + \beta_{1y} \cdot \text{math}_y$$

$$\beta_{0y} = -0.067(0.013) + u_{0y} + \varepsilon_{0y}$$

$$\beta_{1y} = 0.451(0.010) + u_{1y}$$

$$\begin{bmatrix} u_{0y} \\ u_{1y} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.030(0.004) \\ -0.007(0.002) & 0.008(0.003) \end{bmatrix}$$

$$\begin{bmatrix} \varepsilon_{0y} \\ \sigma \end{bmatrix} \sim N(0, \Omega_\varepsilon) : \Omega_\varepsilon = \begin{bmatrix} 0.707(0.009) \end{bmatrix}$$

$$-2 \cdot \log \text{likelihood (ICLS Deviance)} = 31208.650 \text{ (12383 of 12383 cases in use)}$$

The coefficient for $\beta_1$ now has a subscript $j$ ($\beta_{1y}$), and an error component at the school level ($u_{ij}$). Also the random effect of the school level is more complicated. It now has two parts: $u_{0y}$ and $u_{1y}$. $u_{0y}$ represents how much each school intercept varies from the total group intercept (0.030), thus indicating the variability of the school intercepts. $u_{1y}$ represents how much each school slope varies from the total group slope (0.008), thus indicating the variability of the school slopes. The covariance term (-.007) tests whether there is a correlation between differences in slopes and intercepts. This new model is presented graphically in Figure 4.3, showing that not
only do the intercepts of the various schools vary, but the schools’ slopes also vary. As such, it is called a random intercepts and random slopes model.

Figure 4.3. A Random Intercepts and Random Slopes Model

As more predictor variables are added to the simple model just described, the equations and the associated random effects become more complex. In the present investigation, whether variables were allowed to vary across levels depended on the hypotheses and research questions being tested. Such details, along with the composition of the models being tested, are outlined within specific studies.

Procedure

Overview

Data were prepared for analysis using SPSS (v.13). Firstly, missing data issues were resolved and a new weight was calculated to ensure the equivalence of the weighted and unweighted sample sizes, as recommended in the PISA documentation (Organisation for Economic Cooperation and Development, 2005a). Subsequently, variables were standardised, quadratic ability and school-average ability measures created, and cross-products calculated. Data were then transferred to MLwiN (v.2.02) for analysis. Calculations were performed on each plausible value separately, resulting in five sets of results for each analysis conducted. These five sets of results were then combined using Excel (2000). Parameter estimates were averaged and standard errors were calculated to accommodate variance between and
within plausible values, using steps outlined by Raudenbush et al. (2005); see Appendix A.

**Missing Data**

Not all students completed the math self-concept scale that was the main outcome variable of the current study. Students who did not complete these items were therefore deleted from further analyses (10,985 students). Additionally, as previously mentioned, small schools (those with 10 participating students or less) were also deleted from further analyses, as these schools were considered to be too small to be included in multilevel analyses. This resulted in 917 schools (8.9% of total schools) being deleted.

Various variables used in the current study also had data missing. The multilevel modelling program (MLwiN) used to analyse data in the current study uses listwise deletion of non-completers in generating results. However, as the percentage of missing data was minimal ($M = 0.63$, $SD = 1.30$, number of variables = 34), it was decided that no action should be taken.

**Weights**

The database documentation (Organisation for Economic Cooperation and Development, 2005b) recommends that sample weights be used for analyses to prevent estimates of population parameters being biased. The PISA database contained three weights, two country weights and one individual student weight. Marsh and Hau (2003) set each county’s effective sample size equal to the number of that country’s cases, prior to weighting. In order to appropriately extend their findings, the same procedure was followed in the current study. Hence, the individual student weight ($W_{FSTUWT}$) and the country weight ($CNTFAC2$) were combined, as recommended in the database documentation. This ensured that countries had “weights according to their sample sizes so that the sum of weights in each country is equal to the number of students in the database” (Organisation for Economic Cooperation and Development, 2005b, p. 325) and resulted in the weighted and unweighted sample sizes being similar. This combined weight was used to calculate all inferential statistics.
**Standardisation**

The five plausible values for ability, math self-concept, the Individualism Scale scores, SES variables, individual differences in learning variables, and individual perceptions of the learning environment variables were standardised \((M = 0, SD = 1)\) across the entire sample (see Aiken & West, 1991; Marsh & Rowe, 1996; Raudenbush & Bryk, 2002). Quadratic ability variables for each of the five plausible values were created using the standardised plausible values. For each plausible value (linear) a school-average ability variable was calculated by averaging each plausible value separately within each school. Neither the quadratic components of ability, nor the school-average ability variables were restandardised, thus keeping both the school-average and the plausible values (linear and quadratic) for math ability in the same metric.

**Interaction Terms**

Cross-products with school-average ability were created for each plausible value (linear and quadratic), SES variables, individual differences in learning variables, and individual perceptions of the learning environment variables (using standardised variables). As with school-average ability, these product terms were not restandardised. Interaction terms with school-average ability were also created for cultural orientation and economic development.

**Variables Used in Analyses**

Table 4.5 summarises the variables used in the analyses for the current study. In all analyses the outcome variable was individual math self-concept. Predictor variables included individual math ability (linear and quadratic), based on the five plausible values, school-average math ability for the five plausible values, SES variables, individual differences in learning variables, perceptions of the learning environment variables, cultural orientation, and economic development. Also included as predictor variables were the interactions of the five school-average ability variables, based on the plausible values, with the individual linear and quadratic ability plausible values, the SES variables, individual differences in learning variables, the individual perceptions of the learning environment variables, cultural orientation and economic development variables.
### Table 4.5. Summary of Variables Used in Analyses

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Main Effect Predictors</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Self-concept</td>
<td>Math Ability (linear) Plausible Values 1, 2, 3, 4, 5</td>
<td>Math Ability (linear) Plausible Values 1, 2, 3, 4, 5 X School-average Math Ability for Plausible Values 1, 2, 3, 4, 5</td>
</tr>
<tr>
<td></td>
<td>School-average Math Ability for Plausible Values 1, 2, 3, 4, 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Math Ability (quadratic) Plausible Values 1, 2, 3, 4, 5</td>
<td>Math Ability (quadratic) Plausible Values 1, 2, 3, 4, 5 X School-average Math Ability for Plausible Values 1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>SES Variables</td>
<td>SES Variables X School-average Math Ability for Plausible Values 1, 2, 3, 4, 5</td>
<td></td>
</tr>
<tr>
<td>Individual Differences in Learning Variables</td>
<td>Individual Differences in Learning Variables X School-average Math Ability for Plausible Values 1, 2, 3, 4, 5</td>
<td></td>
</tr>
<tr>
<td>Individual Perceptions of the Learning Environment Variables</td>
<td>Individual Perceptions of the Learning Environment Variables X School-average Math Ability for Plausible Values 1, 2, 3, 4, 5</td>
<td></td>
</tr>
<tr>
<td>Cultural Orientation</td>
<td>Cultural Orientation X School-average Math Ability for Plausible Values 1, 2, 3, 4, 5</td>
<td></td>
</tr>
<tr>
<td>Economic Development</td>
<td>Economic Development X School-average Math Ability for Plausible Values 1, 2, 3, 4, 5</td>
<td></td>
</tr>
</tbody>
</table>

### Multilevel Modelling as Used in the Current Study

The PISA 2003 data (see Organisation for Economic Cooperation and Development, 2005a, 2005b) have a three-level hierarchical structure: countries are at the top of the hierarchy, with schools next, and individual students at the bottom of the hierarchy. As noted previously in this chapter, ignoring a hierarchical structure such as this, where students are nested within schools within countries, can result in serious statistical problems. Traditional single level techniques that ignore such a multilevel structure are likely to be invalid as they not only violate assumptions of independence, but they also increase the likelihood of finding statistical significance where none exists (Hox, 2002). Consequently, to accommodate this hierarchical
structure, a multilevel modelling program (MLwiN) was used to analyse these data (refer Rasbash et al., 2002 for details).

A multilevel regression equation consists of two parts: a fixed component and a random component (also called fixed effects and random effects). In the present study, the combination of these fixed and random effects and the treatment of fixed effects were different, depending on the hypothesis or research question being tested. Differences in the treatment of fixed effects were especially evident when the external validity of the BFLPE was investigated, whereby tests were conducted at a cross-national level and at an individual country level.

At the cross-national level, to be consistent with Marsh and Hau (2003), the fixed component consisted of individual ability (both linear and quadratic), school-average ability, their cross-products, and a constant. The random part consisted of the intercepts, variances, and covariances of the levels used in the model. In this case, the intercepts of the country (Level 3), school (Level 2), and student levels (Level 1) were of particular interest, as they showed how much the intercepts of the regression equations varied between countries, between schools, and between students, respectively. Additionally, multilevel modelling can take into account that regression equations of nested data can vary, not only in their intercepts, but also in their slopes. Thus, individual ability (linear and quadratic) was allowed to vary at the school and country level, signifying the extent of variation of the effect of individual ability on math self-concept that existed between countries and between schools. Similarly, school-average ability was allowed to vary at the country level, signifying the extent of variation of the effect of school-average ability on math self-concept that exists between countries.

When the external validity of the BFLPE was examined at the individual country level, the fixed components were individual ability, school-average ability, their cross-products, and a constant. The random components were the intercepts and variances of the two levels used in these models: Level 1 (student) and Level 2 (school). With the exception of the constant (which represents the intercepts) none of the fixed components was allowed to vary, as this was not necessary to test the hypotheses.
Analyses were also conducted at the cross-national level to determine if any of the constructs identified could moderate the BFLPE. The moderators investigated were: cultural orientation, economic development, SES (consisting of five variables), individual differences in learning (consisting of nine variables), and individual perceptions of the learning environment (consisting of three variables). These moderators were examined in separate analyses. For each moderator analysis the fixed components were individual ability (linear), the specific moderator, school-average ability, the cross-product of the moderator and school-average ability, and a constant. The random components for these analyses consisted of the intercepts and variances of the country (Level 3), school (Level 2), and student levels (Level 1). Once again, none of the fixed components, except for the constant, was allowed to vary.

In multiple regression, assumptions of normality and linearity are examined by inspecting residuals. In multilevel regression analysis, normality and linearity are also assumed and can be examined in a similar way by inspecting residuals at different levels (Hox, 2002). In the current study, the assumptions of linearity and normality were assessed for every model specified. No severe violations of these assumptions were noted.

**Combining Results from Analyses of Plausible Values**

Each of the analyses described above was conducted five times, once for each plausible value. When analyses were completed, parameter estimates were averaged and standard errors were calculated using Excel (2000) to accommodate variance between and within plausible values. This was accomplished using the steps outlined by Raudenbush et al. (2005); see Appendix A for details of these steps; see also Organisation for Economic Cooperation and Development (2005a).

**Section Summary**

The preceding section provided a detailed account of the methodology used in Study 1. It included details of participants, measures, and a description of multilevel modelling, the statistical tool used in all three studies of the present investigation. The section concluded by outlining the procedure followed to execute this study,
including details concerning weighting, standardisation, and the use of plausible values. In the following section, the results of analyses are presented. Results for hypotheses and research questions are grouped under the aim they were designed to satisfy.

**Results**

*The External Validity and Generalisability of the BFLPE*

**Overview of Analyses**

To test the external validity and generalisability of the BFLPE, three sets of multilevel regression analyses were conducted. For each set of analyses, results and conclusions are presented for each hypothesis and research question. Results for each set of analyses are then summarised.

The first set of analyses was a test of the external validity of the BFLPE across the entire cross-national sample. Math self-concept was the outcome variable and individual ability (linear and quadratic), school-average math ability, and the interactions of school-average ability with linear and quadratic ability were the predictor variables. These analyses provided results for Hypotheses 1.1, 1.2, and 1.3. (In addition, these analyses were utilised to explore Research Question 2.1; see below).

The second set of analyses tested the external validity of the BFLPE in the 41 individual countries that comprised the cross-national sample. To be consistent with Marsh and Hau (2003), math self-concept was the outcome variable and individual ability (linear and quadratic) and school-average math ability were the predictor variables. These analyses provided results for Hypotheses 1.4 and 1.5.

The third set of analyses tested the moderating effect of cultural orientation and economic development on the BFLPE, using the entire cross-national sample. For each of these moderating analyses, math self-concept was the outcome variable and individual ability (linear), school-average math ability, the moderator (either cultural orientation or economic development), and the interaction of school-average ability with the respective moderator were the predictor variables. These analyses provided results for Research Questions 1.1 and 1.2.
Results for Hypothesis 1.1: Relation Between Individual Math Ability and Math Self-Concept in a Cross-National Sample

Hypothesis 1.1 predicted that there would be a statistically significant positive relation between individual math ability (linear and quadratic) and math self-concept across the cross-national sample of 41 countries. Results indicated that both linear and quadratic components of individual math ability were statistically significantly positively associated with math self-concept, being .52 and .118 respectively (see Table 4.6). Given that the relation between individual math ability (linear and quadratic) and math self-concept was significantly positive across the cross-national sample of 41 countries, Hypothesis 1.1 is accepted.

Results for the linear component of individual ability signified that students whose ability levels were one standard deviation above the mean had math self-concepts that were .52 standard deviations above the average math self-concept score. There was also a positive quadratic relation between individual ability and math self-concept, the nature of which is depicted in Figure 4.4. Controlling for school-average ability, as ability increased so did math self-concept. However, this increase was much more gradual for those at the lower end of the ability spectrum. Compared to students with above average ability, ratings of math self-concept did not increase as steadily for those students whose ability levels were below average.

Results for Hypothesis 1.2: Relation Between Effect of School-Average Math Ability and Math Self-Concept in a Cross-National Sample

Hypothesis 1.2 predicted that there would be a statistically significant negative relation between school-average math ability and math self-concept across the cross-national sample of 41 countries. As indicated in Table 4.6, the relation between school-average math ability and math self-concept was significantly negative (-.30). As school-average math ability had a significantly negative association with math self-concept across the cross-national sample of 41 countries, Hypothesis 1.2 is accepted.

Results indicated that students in schools with an average ability level one standard deviation above the mean had math self-concepts that were .3 of a standard deviation below the average self-concept level. Hence, the negative association
between attending a high-ability school and math self-concept was apparent in a cross-national sample consisting of individualist countries, economically developed countries, economically developing countries, and collectivist countries.

Results for Hypothesis 1.3: Country Variation in the BFLPE

This hypothesis predicted that there would be a statistically significant variation in the relation between school-average math ability and math self-concept across the cross-national sample of 41 countries. Analyses demonstrated (see Table 4.6) a statistically significant, but small, random effect of school-average math ability at the country level (.016), indicating that the size of the BFLPE varied across countries. Given that the BFLPE varied significantly from country to country, Hypothesis 1.3 is accepted.

Summary for Hypotheses 1.1, 1.2, and 1.3

Results for Hypotheses 1.1 and 1.2 indicated that a BFLPE was evident in the cross-national sample. Across the entire sample of 41 countries, higher ability students had higher math self-concepts, but controlling for individual ability, the negative effect of school-average ability indicated that students in high-ability schools had lower math self-concepts than equally able students in low-ability schools.

Although results for Hypothesis 1.3 indicated that the size of the country variation was small, it was statistically significant, indicating that the size of the BFLPE was not consistent across countries. This inconsistency could imply that some countries may have large BFLPEs, while others have none. In order to determine whether individual collectivist and developing countries display a BFLPE, it was necessary to evaluate the BFLPE on a country-by-country basis. This was the focus of Hypotheses 1.4 and 1.5.
### Table 4.6. The Effects of Individual Ability and School Average Ability on Math Self-Concept, in a Cross-National Sample

<table>
<thead>
<tr>
<th>Effect</th>
<th>β</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual ability (linear)</td>
<td>0.520*</td>
<td>0.020</td>
</tr>
<tr>
<td>Individual ability (quadratic)</td>
<td>0.118*</td>
<td>0.005</td>
</tr>
<tr>
<td>School-average ability</td>
<td>-0.300*</td>
<td>0.023</td>
</tr>
<tr>
<td>Interaction effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear ability X school-average ability</td>
<td>-0.048*</td>
<td>0.008</td>
</tr>
<tr>
<td>Quadratic ability X school-average ability</td>
<td>-0.016*</td>
<td>0.005</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.124*</td>
<td>0.036</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3 country intercept</td>
<td>0.052*</td>
<td>0.015</td>
</tr>
<tr>
<td>Level 3 individual ability (linear)</td>
<td>0.014*</td>
<td>0.003</td>
</tr>
<tr>
<td>Level 3 individual ability (quadratic)</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Level 3 school-average ability</td>
<td>0.016*</td>
<td>0.006</td>
</tr>
<tr>
<td>Level 2 School intercept</td>
<td>0.032*</td>
<td>0.005</td>
</tr>
<tr>
<td>Level 2 individual ability (linear)</td>
<td>0.007*</td>
<td>0.003</td>
</tr>
<tr>
<td>Level 2 individual ability (quadratic)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Level 1 Student intercept</td>
<td>0.777*</td>
<td>0.035</td>
</tr>
</tbody>
</table>

*Note. Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SE). Although the \( \beta \) coefficients from a multilevel model are unstandardised, in the present investigation they are equivalent to standardised betas as all variables were standardised before analyses were performed.*
Results for Hypothesis 1.4: Relation Between Individual Math Ability and Math Self-Concept in Individual Countries

Hypothesis 1.4 proposed that there would be a statistically significant positive relation between individual math ability (both linear and quadratic components) and math self-concept in all the individual countries tested. Table 4.7 presents the results of multilevel modelling analyses based on individual countries. In all 41 countries, individual ability (linear) was a significantly positive predictor of math self-concept, ranging from .223 in Indonesia to .786 in Tunisia. Hence, consistent with Marsh and Hau (2003), students with higher math ability tended to have higher math self-concepts. Even in Indonesia, where the effect was smallest, students whose ability levels were one standard deviation above the mean had math self-concepts that were .223 standard deviations above the average math self-concept score.

The quadratic component of individual ability was a significantly positive predictor of math self-concept in all countries except Japan, where its effect was not statistically significant. The effect of quadratic individual ability on math self-concept ranged from .033 in Japan to .175 in Poland, indicating a non-linear relation.
between math ability and math self-concept, as depicted in Figure 4.4 for the entire sample.

Given these results, Hypothesis 1.4 is accepted for the linear component of individual math ability, as its relation with math self-concept was significantly positive in all 41 countries. Additionally, although the effect of the quadratic component of individual ability was not significant in Japan, its relation with math self-concept was significantly positive in 40 of the 41 countries, and was never negative. Hence, Hypothesis 1.4 is accepted for the quadratic component of individual ability, for all countries with the exception of Japan.

Results for Hypothesis 1.5: Relation Between School-Average Math Ability and Math Self-Concept in Individual Countries

Hypothesis 1.5 predicted that, controlling for individual math ability, there would be a statistically significant negative relation between school-average math ability and math self-concept in each individual country examined. As expected from the random effects in the cross-national sample analysis, the size of the BFLPE varied across countries. Results demonstrated that school-average math ability negatively predicted math self-concept in all 41 countries, ranging from -.014 in Korea, to -.713 in Germany (see Table 4.7). This effect was significantly negative in 38 of these countries, and not statistically significant in three (Iceland, Ireland, and Korea). Although the effect of school-average math ability was not significant in three countries, it was always negative and was significantly negative in 38 of the 41 countries. Given these results, Hypothesis 1.5 is accepted for all countries with the exception of Iceland, Ireland, and Korea.

Controlling for individual ability, students in higher ability schools in 38 of the 41 countries had lower math self-concepts than equally able students in lower ability schools. The smallest significant negative effect was in Tunisia where students in schools with an average ability level one standard deviation above the mean had math self-concepts that were .161 of a standard deviation below the average self-concept level. However, in Germany, where the effect was largest, students in schools with an average ability level one standard deviation above the mean had math self-concepts that were .713 of a standard deviation below the average self-concept level.
Non-significant results for the relation between school-average math ability and math self-concept (the BFLPE) were spread among both individualist and collectivist countries (Iceland, Ireland, and Korea). Whereas Ireland scored high on the Individualism Index (70), indicating that it is more of an individualist country, Korea’s score was low (18), indicating that its culture could be regarded as more collectivist in nature (Iceland did not receive a score on this scale.) The World Bank (2007) regards all three countries as high-income (developed) countries. It appears that the BFLPE is just as prevalent in economically developing countries and collectivist countries as it is in developed countries and individualist countries.

Summary for Hypotheses 1.4 and 1.5

Results for Hypotheses 1.4 and 1.5 indicated that a BFLPE was evident in 38 of the 41 countries. In all 41 countries, higher ability students had higher math self-concepts, and controlling for individual ability, in 38 countries the negative effect of school-average math ability indicated that students in high-ability schools had lower math self-concepts than equally able students in low-ability schools. The three countries in which there was not a significant negative effect for school-average math ability were all economically developed countries and were a mixture of individualist and collectivist cultures. Consequently, the results for Hypotheses 1.4 and 1.5 generally indicated that the negative effects of the BFLPE were evident in economically developing, developed, individualist, and collectivist countries.
Table 4.7. The Effects of Individual Ability and School Average Ability on Math Self-Concept, in Individual Countries (Standard Error (SE) in Parentheses)

<table>
<thead>
<tr>
<th>Country (n)</th>
<th>Linear Ability (SE)</th>
<th>Quadratic Ability (SE)</th>
<th>School-Average Ability (SE)</th>
<th>Constant (SE)</th>
<th>School Level (SE)</th>
<th>Individual Level (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia (12383)</td>
<td>0.434* (0.014)</td>
<td>0.077* (0.008)</td>
<td>-0.281* (0.025)</td>
<td>-0.028* (0.014)</td>
<td>0.015* (0.003)</td>
<td>0.710* (0.011)</td>
</tr>
<tr>
<td>Austria (4378)</td>
<td>0.599* (0.030)</td>
<td>0.103* (0.017)</td>
<td>-0.483* (0.053)</td>
<td>-0.076* (0.033)</td>
<td>0.067* (0.010)</td>
<td>0.984* (0.025)</td>
</tr>
<tr>
<td>Belgium (8460)</td>
<td>0.382* (0.019)</td>
<td>0.081* (0.010)</td>
<td>-0.447* (0.028)</td>
<td>-0.120* (0.022)</td>
<td>0.025* (0.005)</td>
<td>0.886* (0.017)</td>
</tr>
<tr>
<td>Brazil (4120)</td>
<td>0.538* (0.036)</td>
<td>0.072* (0.013)</td>
<td>-0.372* (0.039)</td>
<td>0.052 (0.041)</td>
<td>0.025* (0.008)</td>
<td>0.775* (0.018)</td>
</tr>
<tr>
<td>Canada (25685)</td>
<td>0.624* (0.012)</td>
<td>0.118* (0.008)</td>
<td>-0.427* (0.025)</td>
<td>-0.008 (0.014)</td>
<td>0.030* (0.003)</td>
<td>0.942* (0.010)</td>
</tr>
<tr>
<td>Czech Republic (6019)</td>
<td>0.576* (0.021)</td>
<td>0.068* (0.012)</td>
<td>-0.446* (0.029)</td>
<td>-0.245* (0.019)</td>
<td>0.021* (0.005)</td>
<td>0.698* (0.015)</td>
</tr>
<tr>
<td>Denmark (4041)</td>
<td>0.657* (0.021)</td>
<td>0.078* (0.017)</td>
<td>-0.296* (0.055)</td>
<td>0.044 (0.027)</td>
<td>0.023* (0.007)</td>
<td>0.778* (0.017)</td>
</tr>
<tr>
<td>Finland (5730)</td>
<td>0.643* (0.022)</td>
<td>0.149* (0.014)</td>
<td>-0.301* (0.065)</td>
<td>-0.337* (0.038)</td>
<td>0.017* (0.004)</td>
<td>0.762* (0.017)</td>
</tr>
<tr>
<td>France (4154)</td>
<td>0.541* (0.034)</td>
<td>0.111* (0.021)</td>
<td>-0.383* (0.047)</td>
<td>-0.343* (0.031)</td>
<td>0.047* (0.009)</td>
<td>0.922* (0.023)</td>
</tr>
<tr>
<td>Germany (4357)</td>
<td>0.689* (0.030)</td>
<td>0.090* (0.018)</td>
<td>-0.713* (0.042)</td>
<td>0.047 (0.027)</td>
<td>0.032* (0.008)</td>
<td>1.136* (0.024)</td>
</tr>
</tbody>
</table>
Table 4.7. The Effects of Individual Ability and School Average Ability on Math Self-Concept, in Individual Countries (Standard Error (SE) in Parentheses) (contd.)

<table>
<thead>
<tr>
<th>Country (n)</th>
<th>Linear Ability (SE)</th>
<th>Quadratic Ability (SE)</th>
<th>School-Average Ability (SE)</th>
<th>Constant (SE)</th>
<th>School Level (SE)</th>
<th>Individual Level (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greece (4228)</td>
<td>0.523* (0.023)</td>
<td>0.071* (0.013)</td>
<td>-0.174* (0.047)</td>
<td>0.160* (0.022)</td>
<td>0.016* (0.004)</td>
<td>0.694* (0.019)</td>
</tr>
<tr>
<td>Hong Kong (China) (4463)</td>
<td>0.369* (0.023)</td>
<td>0.084* (0.013)</td>
<td>-0.200* (0.035)</td>
<td>-0.517* (0.026)</td>
<td>0.014* (0.004)</td>
<td>0.747* (0.018)</td>
</tr>
<tr>
<td>Hungary (4356)</td>
<td>0.415* (0.025)</td>
<td>0.095* (0.016)</td>
<td>-0.323* (0.042)</td>
<td>-0.268* (0.025)</td>
<td>0.029* (0.006)</td>
<td>0.665* (0.018)</td>
</tr>
<tr>
<td>Iceland (3043)</td>
<td>0.671* (0.030)</td>
<td>0.150* (0.019)</td>
<td>-0.209 (0.125)</td>
<td>-0.265* (0.046)</td>
<td>0.027* (0.009)</td>
<td>1.041* (0.030)</td>
</tr>
<tr>
<td>Indonesia (10448)</td>
<td>0.223* (0.027)</td>
<td>0.063* (0.011)</td>
<td>-0.235* (0.030)</td>
<td>-0.042 (0.032)</td>
<td>0.037* (0.004)</td>
<td>0.374* (0.009)</td>
</tr>
<tr>
<td>Ireland (3801)</td>
<td>0.434* (0.025)</td>
<td>0.120* (0.019)</td>
<td>-0.103 (0.057)</td>
<td>-0.205 (0.023)</td>
<td>0.020* (0.006)</td>
<td>0.758* (0.022)</td>
</tr>
<tr>
<td>Italy (11436)</td>
<td>0.568* (0.016)</td>
<td>0.073* (0.010)</td>
<td>-0.409* (0.028)</td>
<td>-0.122* (0.018)</td>
<td>0.049* (0.006)</td>
<td>0.926* (0.015)</td>
</tr>
<tr>
<td>Japan (4681)</td>
<td>0.355* (0.028)</td>
<td>0.033 (0.017)</td>
<td>-0.307* (0.040)</td>
<td>-0.641 (0.026)</td>
<td>0.029* (0.007)</td>
<td>0.896* (0.022)</td>
</tr>
<tr>
<td>Korea (5380)</td>
<td>0.410* (0.025)</td>
<td>0.085* (0.015)</td>
<td>-0.014 (0.033)</td>
<td>-0.711* (0.024)</td>
<td>0.016* (0.004)</td>
<td>0.669* (0.014)</td>
</tr>
<tr>
<td>Latvia (4522)</td>
<td>0.463* (0.017)</td>
<td>0.121* (0.014)</td>
<td>-0.221* (0.048)</td>
<td>-0.240* (0.020)</td>
<td>0.018* (0.005)</td>
<td>0.556* (0.015)</td>
</tr>
<tr>
<td>Liechtenstein (320)</td>
<td>0.492* (0.076)</td>
<td>0.129* (0.038)</td>
<td>-0.554* (0.061)</td>
<td>0.001 (0.063)</td>
<td>0.000 (0.000)</td>
<td>0.974* (0.125)</td>
</tr>
</tbody>
</table>
Table 4.7.  *The Effects of Individual Ability and School Average Ability on Math Self-Concept, in Individual Countries (Standard Error (SE) in Parentheses)* (contd.)

<table>
<thead>
<tr>
<th>Country (n)</th>
<th>Linear Ability (SE)</th>
<th>Quadratic Ability (SE)</th>
<th>School-Average Ability (SE)</th>
<th>Constant (SE)</th>
<th>School Level (SE)</th>
<th>Individual Level (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxembourg (3870)</td>
<td>0.427* (0.040)</td>
<td>0.130* (0.023)</td>
<td>-0.428* (0.075)</td>
<td>-0.060 (0.036)</td>
<td>0.014* (0.006)</td>
<td>1.191* (0.035)</td>
</tr>
<tr>
<td>Macao(China) (1250)</td>
<td>0.384* (0.037)</td>
<td>0.105* (0.025)</td>
<td>-0.330* (0.091)</td>
<td>-0.352* (0.050)</td>
<td>0.032* (0.013)</td>
<td>0.779* (0.030)</td>
</tr>
<tr>
<td>Mexico (28408)</td>
<td>0.667* (0.014)</td>
<td>0.153* (0.007)</td>
<td>-0.357* (0.019)</td>
<td>0.172* (0.014)</td>
<td>0.029* (0.002)</td>
<td>0.621* (0.007)</td>
</tr>
<tr>
<td>Netherlands (3860)</td>
<td>0.607* (0.046)</td>
<td>0.083* (0.021)</td>
<td>-0.696* (0.051)</td>
<td>-0.083* (0.035)</td>
<td>0.041* (0.009)</td>
<td>0.868* (0.024)</td>
</tr>
<tr>
<td>New Zealand (4428)</td>
<td>0.394* (0.021)</td>
<td>0.114* (0.013)</td>
<td>-0.314* (0.042)</td>
<td>-0.021 (0.027)</td>
<td>0.019* (0.005)</td>
<td>0.670* (0.017)</td>
</tr>
<tr>
<td>Norway (3904)</td>
<td>0.729* (0.022)</td>
<td>0.184* (0.016)</td>
<td>-0.168* (0.069)</td>
<td>-0.419* (0.024)</td>
<td>0.019* (0.006)</td>
<td>0.862* (0.023)</td>
</tr>
<tr>
<td>Poland (4312)</td>
<td>0.534* (0.017)</td>
<td>0.175* (0.014)</td>
<td>-0.279* (0.046)</td>
<td>-0.146 (0.018)</td>
<td>0.013* (0.004)</td>
<td>0.644* (0.018)</td>
</tr>
<tr>
<td>Portugal (4512)</td>
<td>0.579* (0.021)</td>
<td>0.139* (0.017)</td>
<td>-0.205* (0.045)</td>
<td>-0.259* (0.017)</td>
<td>0.010* (0.005)</td>
<td>0.758* (0.019)</td>
</tr>
<tr>
<td>Russian Federation (5822)</td>
<td>0.360* (0.016)</td>
<td>0.072* (0.011)</td>
<td>-0.187* (0.033)</td>
<td>0.069* (0.018)</td>
<td>0.024* (0.005)</td>
<td>0.540* (0.013)</td>
</tr>
<tr>
<td>Serbia and Montenegro (4164)</td>
<td>0.473* (0.022)</td>
<td>0.092* (0.017)</td>
<td>-0.181* (0.043)</td>
<td>0.047 (0.026)</td>
<td>0.021 (0.005)</td>
<td>0.719* (0.020)</td>
</tr>
<tr>
<td>Country (n)</td>
<td>Linear Ability (SE)</td>
<td>Quadratic Ability (SE)</td>
<td>School-Average Ability (SE)</td>
<td>Constant (SE)</td>
<td>School Level (SE)</td>
<td>Individual Level (SE)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>----------------------------</td>
<td>---------------</td>
<td>------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Slovak Republic (7118)</td>
<td>0.550* (0.019)</td>
<td>0.079* (0.013)</td>
<td>-0.411* (0.033)</td>
<td>-0.174 (0.017)</td>
<td>0.029 (0.004)</td>
<td>0.565* (0.012)</td>
</tr>
<tr>
<td>Spain (10695)</td>
<td>0.517* (0.015)</td>
<td>0.119* (0.012)</td>
<td>-0.244* (0.044)</td>
<td>-0.325 (0.018)</td>
<td>0.048* (0.006)</td>
<td>0.857* (0.013)</td>
</tr>
<tr>
<td>Sweden (4513)</td>
<td>0.548* (0.018)</td>
<td>0.137* (0.013)</td>
<td>-0.202* (0.055)</td>
<td>-0.098* (0.023)</td>
<td>0.020* (0.005)</td>
<td>0.756* (0.017)</td>
</tr>
<tr>
<td>Switzerland (8023)</td>
<td>0.436* (0.019)</td>
<td>0.105* (0.012)</td>
<td>-0.446* (0.033)</td>
<td>-0.028* (0.018)</td>
<td>0.031* (0.005)</td>
<td>1.004* (0.018)</td>
</tr>
<tr>
<td>Thailand (5120)</td>
<td>0.276* (0.023)</td>
<td>0.093* (0.012)</td>
<td>-0.194* (0.035)</td>
<td>-0.163* (0.024)</td>
<td>0.018* (0.005)</td>
<td>0.422* (0.013)</td>
</tr>
<tr>
<td>Tunisia (4653)</td>
<td>0.786* (0.069)</td>
<td>0.138* (0.024)</td>
<td>-0.161* (0.062)</td>
<td>0.586* (0.057)</td>
<td>0.038* (0.008)</td>
<td>1.195* (0.027)</td>
</tr>
<tr>
<td>Turkey (4652)</td>
<td>0.520* (0.032)</td>
<td>0.054* (0.016)</td>
<td>-0.252* (0.040)</td>
<td>-0.063* (0.030)</td>
<td>0.030* (0.006)</td>
<td>0.883* (0.024)</td>
</tr>
<tr>
<td>United Kingdom (9262)</td>
<td>0.457* (0.016)</td>
<td>0.120* (0.012)</td>
<td>-0.344* (0.029)</td>
<td>-0.060* (0.015)</td>
<td>0.020* (0.004)</td>
<td>0.739* (0.013)</td>
</tr>
<tr>
<td>United States (5202)</td>
<td>0.516* (0.018)</td>
<td>0.132* (0.013)</td>
<td>-0.230 (0.043)</td>
<td>0.131* (0.020)</td>
<td>0.023* (0.005)</td>
<td>0.929* (0.019)</td>
</tr>
<tr>
<td>Uruguay (5407)</td>
<td>0.575* (0.026)</td>
<td>0.091* (0.013)</td>
<td>-0.240* (0.039)</td>
<td>0.080* (0.026)</td>
<td>0.041* (0.007)</td>
<td>0.911* (0.019)</td>
</tr>
</tbody>
</table>

*Note.* Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SE).
Results for Research Question 1.1: Culture as a Moderator of the BFLPE

Research Question 1.1 asked whether, using the cross-national sample, the cultural orientation of a country could moderate the BFLPE. In particular, it enquired about the main effect of cultural orientation on math self-concept and about the relation between the school-average math ability X cultural orientation interaction and math self-concept. As seen in Table 4.8, analyses of the cross-national sample exploring culture as a moderator of the BFLPE indicated that a BFLPE was present (main effect of individual ability on math self-concept was significantly positive, .529, and the main effect of school-average ability on math self-concept was significantly negative, -.380). Neither the main effect of cultural rating (.035), an indication of how individualist/collectivist a country is, nor the interaction effect of the cultural rating X school-average math ability interaction (.015) on math self-concept was statistically significant.

These results indicated that the BFLPE was consistent across cultures, as the interaction term (cultural rating X school-average math ability) was not statistically significant. Irrespective of whether a country espoused an individualist or collectivist culture, students in high-ability schools had lower math self-concepts than equally able students in low-ability schools. As such cultural orientation did not moderate the effects of the BFLPE.

Results for Research Question 1.2: Economic Development as a Moderator of the BFLPE

This research question was posed to explore the moderating effect of economic development on the BFLPE in the cross-national sample. In particular, the main effect of the economic development of a country and the interaction effect of the school-average math ability X stage of economic development interaction on math self-concept were examined. Analyses of the cross-national sample exploring economic development as a moderator of the BFLPE indicated that a BFLPE was present (the main effect of individual ability on math self-concept was significantly positive, .528, and the main effect of school-average ability on math self-concept was significantly negative, -.401; see Table 4.8). Results for the main effect of economic development on math self-concept indicated that students in economically developed countries had math self-concepts that were .126 of a standard deviation significantly
lower than of students in developing countries. However, the economic development X school-average math ability interaction (.046) was not statistically significant.

The interaction of economic development and school-average math ability was not statistically significant, indicating that the BFLPE was similar across both economically developing and developed countries. Although students in developed countries tended to have lower math self-concepts than students in developing countries, being in a high-ability class had a similar negative association with math self-concept, irrespective of the economic development of the country. Hence economic development did not moderate the BFLPE.

Summary for Research Questions 1.1 and 1.2

Findings from Research Questions 1.1 and 1.2 provided further support for the external validity and generalisability of the BFLPE. Results of moderating analyses demonstrated that the BFLPE occurs in both individualist and collectivist cultures and in both economically developed and developing nations.

Table 4.8. Cultural and Economic Moderators of the BFLPE
(Standard Error (SE) in Parentheses)

<table>
<thead>
<tr>
<th>Mod Indiv/Collect</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod Ability (SE)</td>
<td>Sch Avg (SE)</td>
<td>Mod Var (SE)</td>
</tr>
<tr>
<td>Individ</td>
<td>0.529*</td>
<td>-0.380*</td>
</tr>
<tr>
<td>/Collect</td>
<td>(0.030)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Econ Develop a</td>
<td>0.528*</td>
<td>-0.401*</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.035)</td>
</tr>
</tbody>
</table>

Note. Mod = Moderator; Individ/Collect = Individualist/Collectivist; Econ Develop = Economic Development; Ind Ability = Individual ability; Mod Var = Moderator variable; Sch Avg = School-average ability; Mod X Sch Avg = Moderator by school-average interaction; Cntry = Country; Cons = Constant. Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SE). a = Developing countries constitute the reference group.

Section Summary: The External Validity and Generalisability of the BFLPE

Evidence attesting to the external validity and generalisability of the BFLPE was provided in three sets of analyses. Firstly, the BFLPE was shown to exist in a cross-national sample of 41 countries that comprised economically developed,
economically developing, individualist, and collectivist nations. These countries were then analysed separately, with a BFLPE emerging in 38 of the 41 countries. The three countries not displaying a BFLPE were all economically developed and included both individualist and collectivist nations. The final set of analyses showed that the BFLPE was not moderated by cultural orientation or by economic development. Overall, these results suggest that the BFLPE is externally valid and universally applicable.

**Individual Ability and the BFLPE**

**Overview of Analyses**

The moderating effect of individual ability on the BFLPE was tested as part of the cross-national multilevel regression analyses conducted to investigate the external validity of the BFLPE. As previously, the outcome variable was math self-concept, and predictor variables were individual ability (linear and quadratic), school-average math ability, and the interactions of school-average ability with linear and quadratic ability. These analyses provided results for Research Question 2.1.

**Results for Research Question 2.1: Individual Math Ability as a Moderator of the BFLPE**

This research question explored whether individual math ability (linear and quadratic components) could moderate the BFLPE. Specifically, the relation between math self-concept and the individual math ability X school-average math ability interaction in the cross-national sample of 41 countries was examined. As seen in Table 4.6, the relation between the individual math ability (linear) by school-average math ability interaction and math self-concept was statistically negatively significant (-.048). Results were similar for the quadratic individual math ability by school-average math ability interaction. Although smaller, this interaction was also statistically negatively significant (-.016).

The pattern of the individual ability X school-average interaction is depicted in Figure 4.5. The figure indicates that the BFLPE was somewhat worse for high-ability students attending high-ability schools, as reflected in the somewhat steeper slopes
for students whose ability was one to two standard deviations above the mean. However, students of all ability levels suffered the BFLPE whereby in high-ability schools students of all ability levels had lower math self-concepts than their counterparts in low-ability schools. Nevertheless, although the individual ability by school-average ability interaction was significantly negative, it was small, especially given the power engendered by the large sample size. Moreover, although there was a significantly negative quadratic interaction, this effect was so small that it is not visually evident in the graph of the interaction equation.

**Section Summary: Individual Ability and the BFLPE**

These results suggest that the BFLPE may be greater for high-ability students. Nevertheless, the interaction effects were small. Given the size of the sample and the power of the tests, that these interaction effects were small supports the viewpoint that the BFLPE generalises well over ability levels.

![Figure 4.5](image)

**Figure 4.5.** Math Self-Concept as a Function of School-Average Ability and Individual Ability

*Note.* Based on predicted values. -2SD = two standard deviations below the mean of school-average ability; -1SD = one standard deviation below the mean of school-average ability; +1SD = one standard deviation above the mean of school-average ability; +2SD = two standard deviations above the mean of school-average ability. –2SD Ability = two standard deviations below the mean of individual ability; -1SD Ability = one standard deviation below the mean of individual ability; +1SD Ability = one standard deviation above the mean of individual ability; +2SD Ability = two standard deviations above the mean of individual ability.
Overview of Analyses

A separate multilevel regression analysis was conducted for each potential moderator of the BFLPE (five SES moderators, nine individual differences in learning moderators, and three perceptions of the learning environment moderators), using the cross-national sample. For each potential moderator, results for its main effect are presented, followed by results of its interaction with school-average math ability. Both sets of results are subsequently summarised together. In all analyses the outcome variable was math self-concept, and predictor variables were individual math ability, school-average math ability, the potential moderator, and the interaction of school-average math ability with the relevant moderator variable. These analyses provided results for Hypotheses 3.1 through to 3.7, and for Research Questions 3.1 through to 3.7.

Results for Hypothesis 3.1: Relation Between Individual Socio-Economic Status (SES) and Math Self-Concept

Hypothesis 3.1 proposed that, in the cross-national sample, each of the individual SES variables would have a statistically significant positive association with math self-concept. Although the effect of the highest parental occupation level on math self-concept was not statistically significant (.007), the remaining four SES variables tested had a significantly positive, but small, association with math self-concept (highest in education, .039; home education resources, .038; cultural possessions, .019; and economic, social, and cultural status, .039). These positive main effects indicate that, controlling for individual ability, higher math self-concepts were weakly associated with students from higher SES backgrounds. As four of the SES variables had a significantly positive association with math self-concept, Hypothesis 3.1 is accepted for highest in education, home education resources, cultural possessions, and economic, social, and cultural status, but not accepted for highest in occupation (see Table 4.9).
Results for Research Question 3.1: Individual SES as a Moderator of the BFLPE

This research question asked whether SES could moderate the BFLPE. In particular Research Question 3.1 enquired about the relation in the cross-national sample between math self-concept and the interaction of each of the individual SES variables with school-average math ability.

A BFLPE was evident when all SES variables were tested as moderators of the BFLPE (see Table 4.9). Individual ability was always a positive predictor of math self-concept, ranging from .521 for economic, social, and cultural status to .536 for parent highest in occupation. School-average math ability was a negative predictor of math self-concept in all instances, ranging from -.378 for parent highest in occupation to -.391 for economic, social, and cultural status.

Of critical importance to the current research question was whether the BFLPE was moderated by any of these SES variables. Two interactions with school-average math ability were statistically significant, but small: highest in occupation (.027) and home education resources (.026). The pattern of these interactions is presented in Figures 4.6 and 4.7.

Figure 4.6 indicates that smaller BFLPEs were associated with students whose parents had more prestigious occupations. In low-ability schools students had similar math self-concepts irrespective of their parents’ occupations. However, in high-ability schools, compared to students whose parents held less prestigious occupations, smaller BFLPEs were associated with students whose parents had more prestigious occupations. As regards educational resources in the home, Figure 4.7 indicates that smaller BFLPEs were associated with homes that contained more educational resources. Math self-concepts for students in low-ability schools who had many educational resources in the home were slightly higher compared to those who had less. This difference increased slightly in high-ability schools such that compared to students in homes with fewer education resources, smaller BFLPEs were associated with students whose homes contained more educational resources.

Summary for Hypothesis 3.1 and Research Question 3.2: SES

Results indicated that SES was a significant predictor of math self-concept and that smaller BFLPEs were associated with students whose parents had more
prestigious occupations and for students in homes with more educational resources. However, the sizes of the main effects were small as were the interaction effects, especially given that the large sample size would have greatly enhanced the power of these tests.

In unreported supplemental analyses, the effect of school-average SES was evaluated. In all cases, school-average SES did not add to the prediction of math self-concept over and above that of the individual SES variables. Thus, school-average SES was not considered further.

Table 4.9.  

<table>
<thead>
<tr>
<th>SES Moderators of the BFLPE (Standard Error (SE) in Parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Modifier</td>
</tr>
<tr>
<td>Highest in Occupation</td>
</tr>
<tr>
<td>Highest in Education</td>
</tr>
<tr>
<td>Home Education Resources</td>
</tr>
<tr>
<td>Cultural Possessions</td>
</tr>
<tr>
<td>Economic, Social and Cultural Status</td>
</tr>
</tbody>
</table>

*Note. Ind Ability = Individual Ability; Mod Var = Moderator Variable; Sch Avg = School-average Ability; Mod X Sch Avg = Moderator by School-average Ability; Cntry = Country; Sch = School; Ind = Individual Student; Cons = Constant. Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SE).
Figure 4.6. Higher in Occupation by School-Average Ability Interaction

*Note.* Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability and low school-average ability = 1 standard deviation below the mean for school-average ability. Similarly, high occupation = 1 standard deviation above the mean for occupation, and lower in occupation = 1 standard deviation below the mean. Individual ability is held constant at its mean.

Figure 4.7. Home Education Resources by School-Average Ability Interaction

*Note.* Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability and low school-average ability = 1 standard deviation below the mean. Similarly, higher education resources = 1 standard deviation above the mean for home education resources, and lower education resources = 1 standard deviation below the mean. Individual ability is held constant.
Results for Hypothesis 3.2: Relation Between Self-Regulated Learning Strategies and Math Self-Concept

Hypothesis 3.2 predicted that, in the cross-national sample, there would be a statistically significant positive association between math self-concept and each self-regulated learning strategy (elaboration, memorisation, and control strategies). Results indicated that all three self-regulated learning strategies had a statistically significant positive relation with math self-concept (elaboration = .367, memorisation = .261, and control strategies = .233; see Table 4.10). As results were as predicted, Hypothesis 3.2 is accepted.

Results for Research Question 3.2: Self-Regulated Learning Strategies as Moderators of the BFLPE

Research Question 3.2 posed the question as to whether the BFLPE could be moderated by self-regulated learning strategies. The relation between math self-concept and each of the interactions of the self-regulated learning strategies (elaboration, memorisation, and control strategies) with school-average math ability in the cross-national sample was examined.

A BFLPE was evident when all self-regulated learning strategy variables were tested as moderators of the BFLPE (see Table 4.10). Math self-concept was always positively predicted by individual ability ranging from .496 for elaboration to .524 for memorisation. Math self-concept was negatively predicted by school-average math ability in all instances, ranging from -.305 for elaboration to -.372 for control strategies.

Whereas the interaction of the self-regulated learning strategy, elaboration, with school-average math ability had a significant positive association with math self-concept (.019; see Figure 4.8), that of memorisation had a significant negative association (-.070; see Figure 4.9). The interaction of control strategy use with school-average math ability was not statistically significant (-.021). Slightly smaller negative effects of the BFLPE were associated with students in high-ability schools who used the learning strategy elaboration to a greater degree, although as indicated in Figure 4.8, this difference is barely discernable. Compared to their counterparts in low-ability schools, students in high-ability schools had lower
math self-concepts. However, students in high-ability schools who used elaboration techniques to a greater extent had very slightly higher math self-concepts than those who used this learning strategy to an average or low extent. Nonetheless, as revealed in the graph of the interaction, the slightly smaller BFLPE for those higher in elaboration was minimal.

Figure 4.9 shows that larger BFLPEs were associated with students who used memorisation to a greater extent. Students in high-ability schools had lower math self-concepts than their counterparts in low-ability schools. However, students in high-ability schools who used the memorisation technique to a greater extent suffered a larger decline in math self-concept than those who used the technique to a lesser extent. Hence, greater negative effects of the BFLPE were associated with those who relied more on memorisation as a learning strategy.

Summary for Hypothesis 3.2 and Research Question 3.2: Self-Regulated Learning Strategies

Students who used elaboration, memorisation, and control strategies as techniques to improve their learning had higher math self-concepts. Tests of the moderating effects of these self-regulated learning strategies demonstrated significant interaction effects for memorisation and elaboration. Using elaboration techniques were associated with smaller BFLPEs, but this interaction effect was small, especially when the power engendered by the large sample is considered. The interaction effect for memorisation was larger and suggested that using memorisation techniques were associated with larger BFLPEs.
Table 4.10. *Individual Differences in Learning Moderators of the BFLPE*

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moderator</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Iabil</td>
<td>Savg</td>
</tr>
<tr>
<td>Elaboration</td>
<td>0.496*</td>
<td>-0.305*</td>
</tr>
<tr>
<td>(0.027)</td>
<td>(0.022)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Memorisation</td>
<td>0.524*</td>
<td>-0.351*</td>
</tr>
<tr>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Control Strategies</td>
<td>0.511*</td>
<td>-0.372*</td>
</tr>
<tr>
<td>(0.027)</td>
<td>(0.024)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrinsic</td>
<td>0.442*</td>
<td>-0.286*</td>
</tr>
<tr>
<td>(0.022)</td>
<td>(0.021)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Intrinsic</td>
<td>0.358*</td>
<td>-0.209*</td>
</tr>
<tr>
<td>(0.019)</td>
<td>(0.022)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Self Related Cognitions</td>
<td>Math Self-</td>
<td></td>
</tr>
<tr>
<td>efficacy</td>
<td>0.281*</td>
<td>-0.357*</td>
</tr>
<tr>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Math Anxiety</td>
<td>0.247*</td>
<td>-0.0215*</td>
</tr>
<tr>
<td>(0.014)</td>
<td>(0.016)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Learning Environment</td>
<td>Competitive</td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>0.479*</td>
<td>-0.303*</td>
</tr>
<tr>
<td>(0.025)</td>
<td>(0.023)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Cooperative Learning</td>
<td>0.534*</td>
<td>-0.369*</td>
</tr>
<tr>
<td>(0.029)</td>
<td>(0.026)</td>
<td>(0.019)</td>
</tr>
</tbody>
</table>

Note. Standard Error (SE) in Parentheses; IAbil = Individual Ability; Mod = Moderator Variable; Savg = School-average Ability; ModXSavg = Moderator by School-average Ability; Cntry = Country; Sch = School; Ind = Individual Student; Cons = Constant. Note: Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs).
Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability, and low school-average ability = 1 standard deviation below the mean. Similarly, high elaboration = 1 standard deviation above the mean for elaboration, and low elaboration = 1 standard deviation below the mean. Individual ability is held constant.

**Figure 4.8. Elaboration by School-Average Ability Interaction**

*Note.* Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability, and low school-average ability = 1 standard deviation below the mean. Similarly, high elaboration = 1 standard deviation above the mean for elaboration, and low elaboration = 1 standard deviation below the mean. Individual ability is held constant.

**Figure 4.9. Memorisation by School-Average Ability Interaction**

*Note.* Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability, and low school-average ability = 1 standard deviation below the mean. Similarly, high memorisation = 1 standard deviation above the mean for memorisation, and low memorisation = 1 standard deviation below the mean. Individual ability is held constant.
**Results for Hypothesis 3.3: Relation Between Motivation and Math Self-Concept**

The prediction formulated for this hypothesis was that, in the cross-national sample, there would be a statistically significant positive relation between both intrinsic and extrinsic motivation and math self-concept. Consistent with this prediction, both extrinsic and intrinsic motivation were statistically significant positive predictors of math self-concept (extrinsic = .384; intrinsic = .625; see Table 4.10). Students who were motivated, extrinsically or intrinsically, had higher math self-concepts. Given these results, Hypothesis 3.3 is accepted.

**Results for Research Question 3.3: Motivation as a Moderator of the BFLPE**

This research question asked whether motivation could moderate the BFLPE. In particular, it questioned what the relation would be between math self-concept and the motivation (both intrinsic and extrinsic) X school-average math ability interactions, in the cross-national sample.

Analyses exploring extrinsic and intrinsic motivation as moderators of the BFLPE in the cross-national sample indicated the presence of the BFLPE (see Table 4.10). For both types of motivation, the main effect of individual ability on math self-concept was significantly positive (extrinsic = .442; intrinsic = .358) and the main effect of school-average math ability on math self-concept was significantly negative (extrinsic = -.286; intrinsic = -.209).

The interaction of extrinsic motivation and school-average math ability had a significant, but small, positive association with math self-concept (.040). As depicted in Figure 4.10, slightly smaller BFLPEs were associated with students in high-ability schools who were high in extrinsic motivation. Compared to equally extrinsically motivated students in low-ability schools, students in high-ability schools had lower math self-concepts. However, in high-ability schools, students low in extrinsic motivation displayed a slightly larger drop in math self-concept. Nevertheless, as seen in Figure 4.10, this slightly larger BFLPE for less extrinsically motivated students was negligible.

The intrinsic motivation X school-average math ability interaction also had a significant, but smaller, positive effect (.015). An inspection of the graph of the interaction (see Figure 4.11) tells a similar story to that of extrinsic motivation.
Slightly smaller BFLPEs were associated with students in high-ability schools who were high in intrinsic motivation, although this buffering effect was less than that for extrinsic motivation. Additionally, students in high-ability schools had lower math self-concepts when compared to equally intrinsically motivated students in low-ability schools. However, students low in intrinsic motivation in high-ability schools displayed a slightly larger drop in math self-concept, although once again, this interaction effect is barely discernable in the graph of the interaction.

**Summary for Hypothesis 3.3 and Research Question 3.3: Motivation**

Higher math self-concepts were associated with both extrinsic and intrinsic motivation. However, the size of the intrinsic main effect was larger, with a corresponding reduction in the contribution of both individual math ability and school-average math ability to the prediction of math self-concept. The significant interaction effects suggest that being highly motivated can buffer students against the negative effects of the BFLPE. Nevertheless, when the power produced by the large sample size is taken into account, both interaction effects were trivial.

![Graph showing Extrinsic Motivation by School-Average Ability Interaction](image)

*Figure 4.10. Extrinsic Motivation by School-Average Ability Interaction*

*Note. Based on predicted values. High school-average ability = 1 standard deviation above the mean for school average math ability, and low school-average ability = 1 standard deviation below the mean. Similarly, high extrinsic motivation = 1 standard deviation above the mean for extrinsic motivation, and low instrumental motivation = 1 standard deviation below the mean. Individual ability is held constant.*
Results for Hypothesis 3.4: Relation Between Math Self-Efficacy and Math Self-Concept

Hypothesis 3.4 predicted that there would be a statistically significant positive relation between math self-efficacy and math self-concept in the cross-national sample. As evidenced in Table 4.10, math self-efficacy was a significantly positive predictor of math self-concept (.440). Students who were high in math self-efficacy had higher math self-concepts. As predictions were upheld, Hypothesis 3.4 is accepted.

Results for Research Question 3.4: Math Self-Efficacy as a Moderator of the BFLPE

Research Question 3.4 enquired about the moderating effect of math self-efficacy on the BFLPE. Specifically, the relation between math self-concept and the math self-efficacy X school-average math ability interaction was examined.

When math self-efficacy was tested as a BFLPE moderator, results indicated that a BFLPE was evident (see Table 4.10). Individual ability was a positive predictor of math self-concept (.281) and school-average math ability was a negative predictor (-.357). The math self-efficacy X school-average math ability interaction had a
significantly positive association with math self-concept (.019), and is depicted in Figure 4.12. Compared to students low in self-efficacy, slightly smaller BFLPEs were associated with students high in self-efficacy. However, compared to their counterparts in low-ability schools, all students in high-ability schools had lower math self-concepts, irrespective of their level of math self-efficacy.

Summary for Hypothesis 3.4 and Research Question 3.4: Self-Efficacy

When they felt that they were able to succeed in math, students in the current study felt better about their accomplishments in math. Additionally, students high in self-efficacy were buffered against the negative effects of the BFLPE. Nevertheless, this interaction effect was small and has to be considered in the context of the sample size.

![Graph showing Math Self-Efficacy by School-Average Ability Interaction](image)

**Figure 4.12.** Math Self-Efficacy by School-Average Ability Interaction

*Note.* Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability, and low school-average ability = 1 standard deviation below the mean. Similarly, high math efficacy = 1 standard deviation above the mean for math efficacy, and low math efficacy = 1 standard deviation below the mean. Individual ability is held constant.

Results for Hypothesis 3.5: Relation Between Math Anxiety and Math Self-Concept

This hypothesis proposed that, in the cross-national sample, math anxiety would have a statistically significant negative association with math self-concept. As indicated in Table 4.10, math anxiety had a large statistically significant negative
association with math self-concept (-.599). Highly anxious students had lower math self-concepts. Given these results, Hypothesis 3.5 is accepted.

Results for Research Question 3.5: Math Anxiety as a Moderator of the BFLPE

The focus of this research question was on the moderating effect of math anxiety on the BFLPE. Specifically, it examined the relation between the math anxiety X school-average math ability interaction and math self-concept, in the cross-national sample. Results indicated the presence of the BFLPE. Whereas individual ability was positively related to math self-concept (.247), school-average math ability displayed a negative relation (-.215).

The math anxiety X school-average math ability interaction had a significantly negative association with math self-concept (-.116); see Figure 4.13. High math anxiety was associated with the BFLPE, but low math anxiety was not. Students with low anxiety levels had similar math self-concepts irrespective of the ability level of the school they attended. Conversely, compared to those who had average or low anxiety levels, larger BFLPEs were associated with students in high-ability schools who were highly anxious about their math studies. Students in high-ability schools with average anxiety levels showed a slight decline in their math self-concepts compared to their counterparts in low-ability schools. However, in high-ability schools, students who were highly anxious about math had considerably lower math self-concepts than their counterparts in low-ability schools.

Summary for Hypothesis 3.5 and Research Question 3.5: Anxiety

In the current study, highly anxious students had lower math self-concepts. The significant interaction effect of anxiety and school-average math ability on math self-concept suggests that being low in anxiety may be a protective factor against the negative effects of the BFLPE.
Figure 4.13. Math Anxiety by School-Average Ability Interaction

Note. Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability and low school-average ability = 1 standard deviation below the mean. Similarly, high math anxiety = 1 standard deviation above the mean for math anxiety, and low math anxiety = 1 standard deviation below the mean. Individual ability is held constant.

Results for Hypothesis 3.6: Relation Between Preference for Learning Environment and Math Self-Concept

This hypothesis envisaged that preferences for a competitive or a cooperative learning environment would have statistically significant positive associations with math self-concept in the cross-national sample. These predictions were supported. Results, presented in Table 4.10, indicated that both a competitive and a cooperative preference for the learning environment had a statistically significant positive relation with math self-concept (.368 and .119 respectively), although this effect was stronger for those who preferred a competitive environment. Students who preferred a competitive or a cooperative learning environment had higher math self-concepts. As results were consistent with predictions, Hypothesis 3.6 is accepted.

Results for Research Question 3.6: Preference for Learning Environment as a Moderator of the BFLPE

Research Question 3.6 enquired about the moderating effect of a student’s preference for a learning environment on the BFLPE. It explored the relations between the competitive preference for learning environment X school-average math
ability interaction and the cooperative preference for learning environment X school-average math ability interaction with math self-concept in the cross-national sample.

When preferences for type of learning environment were examined, a BFLPE was present. Individual ability had a significantly positive relation with math self-concept (competitive learning = .479; cooperative learning = .534) and school-average math ability had a negative effect (competitive learning = -.303; cooperative learning = -.369). Whereas the competitive preference for learning environment X school-average math ability interaction failed to reach significance (-.006), the cooperative preference for learning environment X school-average math ability interaction had a statistically significant, but small, negative association with math self-concept (-.084; see Figure 4.14). Compared to those who had a low preference for a cooperative learning environment, larger BFLPEs were associated with students who had a high preference for a cooperative learning environment. Whether they espoused a cooperative preference to a greater or lesser extent, math self-concept was reduced for those students who attended high-ability schools. However, this reduction in math self-concept in high-ability schools was less for those who espoused a low preference for a cooperative learning environment.

**Summary for Hypothesis 3.6 and Research Question 3.6: Preference for Learning Environment**

Preferences for competitive or cooperative learning environments were both associated with higher math self-concepts. Having a preference for a competitive learning environment had no effect on the BFLPE: Students in high-ability schools had lower math self-concepts, irrespective of the extent to which they espoused a preference for a competitive environment. Conversely, having a preference for a cooperative learning environment was associated with larger BFLPEs. Although not a large interaction effect, this finding suggests that students who prefer to learn in a cooperative environment are more prone to the negative effects of the BFLPE.
Results for Hypothesis 3.7: Relation Between Individual Perceptions of the Learning Environment and Math Self-Concept

Hypothesis 3.7 predicted that each of the individual perceptions of the learning environment variables (perceptions of student-teacher relations, a sense of belonging to the school, and attitudes to school) would have a statistically significant positive relation with math self-concept in the cross-national sample. Results demonstrated that perceptions of student-teacher relations (.148), a sense of belonging to the school (.086), and attitudes to school (.112) all displayed a statistically significant positive relation with math self-concept (see Table 4.11). Given that all three variables displayed a positive relation with math self-concept, Hypothesis 3.7 is accepted.

Results for Research Question 3.7: Individual Perceptions of the Learning Environment as Moderators of the BFLPE

The extent to which individual perceptions of the learning environment moderated the BFLPE was examined in this research question. It examined the relation between math self-concept and the individual perceptions of the learning environment (perceptions of student-teacher relations, a sense of belonging to the
school, and attitudes to school) X school-average math ability interactions, in the cross-national sample.

The BFLPE was evident in all these analyses as individual ability had a statistically significant positive association with math self-concept ranging from .521 (attitude to school) to .528 (sense of belonging), and school-average math ability had a statistically significant negative association with math self-concept, ranging from -.364 (student-teacher relations) to -.384 (sense of belonging); see Table 4.11.

The student-teacher relations X school-average math ability interaction was not statistically significant (.004). Irrespective of their perceptions of student-teacher relations in their schools, students in high-ability schools had lower math self-concepts. Statistically significant positive interactions with school-average math ability were demonstrated for sense of belonging (.017) and attitudes to school (.048). Students in high-ability schools had lower math self-concepts than their counterparts in low-ability schools, but this effect was somewhat worse for students who felt less of a sense of belonging to their school. Students in high-ability schools who felt more of a sense of belonging to their school did not experience as large a drop in their math self-concepts as those who felt less of a sense of belonging, although this effect is barely discernable in the graph of the interaction (see Figure 4.15). Similarly, students in high-ability schools who had more positive attitudes to school suffered less from the BFLPE (see Figure 4.16). Although students in high-ability schools had lower math self-concepts than their counterparts in low-ability schools, those who perceived school more positively were less affected by attending a high-ability school than those who held less positive attitudes.

Summary for Hypothesis 3.7 and Research Question 3.7: Individual Perceptions of the Learning Environment

In the current study, more positive math self-concepts were associated with individual perceptions of a sense of belonging to the school, more positive attitudes to school, and good student-teacher relationships. The BFLPE was smaller when students held more positive attitudes to school or when they felt a greater sense of belonging to the school. Once again, these interactions were small, especially when the size of the sample is taken into consideration.
Table 4.11. *Individual Perceptions of the Learning Environment as Moderators of the BFLPE (Standard Error (SE) in Parentheses)*

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ind Ability</td>
<td>Sch Avg</td>
</tr>
<tr>
<td>Student-Teacher Relations</td>
<td>0.525*</td>
<td>-0.364*</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Sense of Belonging</td>
<td>0.528*</td>
<td>-0.384*</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Attitude to School</td>
<td>0.521*</td>
<td>-0.370*</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.024)</td>
</tr>
</tbody>
</table>

*Note.* Ind Ability = Individual Ability; Mod Var = Moderator Variable; Sch Avg = School-average Ability; Mod X Sch Avg = Moderator by School-average Ability; Cntry = Country; Sch = School; Ind = Individual Student; Cons = Constant. Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SE).
Note. Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability, and low school-average = 1 standard deviation below the mean. Similarly, high sense of belonging = 1 standard deviation above the mean for sense of belonging and low sense of belonging = 1 standard deviation below the mean. Individual ability is held constant.

Figure 4.15. Sense of Belonging by School-Average Ability Interaction

Note. Based on predicted values. High school-average ability = 1 standard deviation above the mean for school-average math ability, and low school-average = 1 standard deviation below the mean. Similarly, more positive attitudes to school = 1 standard deviation above the mean for attitude to school, and less positive attitudes to school = 1 standard deviation below the mean. Individual ability is held constant.

Figure 4.16. Attitude to School by School-Average Ability Interaction
Section Summary: Individual Student Characteristics and the BFLPE: SES, Individual Differences in Learning, and Perceptions of the Learning Environment

Higher math self-concepts were associated with higher SES, various learning characteristics, and perceptions of the learning environment. The use of learning techniques (such as elaboration, memorisation, and control strategies), being motivated either extrinsically or intrinsically, having a feeling of self-efficacy and a preference for either a competitive or a cooperative learning environment all had a statistically significant positive effect on math self-concept. Students who felt they belonged to the school, had more positive attitudes to school, and considered student-teacher relationships to be good, had higher math self-concepts. Conversely, students who were highly anxious about their math studies had lower math self-concepts.

Statistically significant interactions with school-average math ability demonstrated that the BFLPE was moderated by many of the constructs tested. The BFLPE was smaller for students who had parents with more prestigious occupations; had more educational resources in their homes; used elaboration as a learning strategy; were motivated both intrinsically and extrinsically; felt able to succeed; had more positive attitudes to school; and felt a greater sense of belonging to their school. Conversely, the BFLPE was larger for students who were highly anxious, who used memorisation as a learning strategy, and who preferred to learn in a cooperative environment.

However, there are caveats that apply to these results. Although interactions for anxiety, memorisation, and a preference for cooperative learning were acceptable, many of the statistically significant interactions with school-average math ability were small and when examined graphically appeared to be trivial. Interactions for elaboration, extrinsic motivation, intrinsic motivation, and math self-efficacy revealed very little differences between different levels of these factors. Graphs of the interactions of occupational status, home educational resources, and attitudes to school displayed slightly larger differences between levels, indicating that these factors might have somewhat of a buffering effect on the BFLPE. Nevertheless, considering the power engendered by the large sample size, results for these moderators should be treated cautiously.


Section Summary

Results of analyses conducted to test the hypotheses and research questions pertaining to Study 1 were presented in this section. Results were grouped according to the particular aim that hypotheses and research questions were designed to satisfy. Each sub-section began with an overview of the analyses used, continued with a presentation of findings, and concluded with a summary of the results.

Discussion

Overview

The current study aimed to: (a) test the external validity and generalisability of the BFLPE; (b) test whether the size of the BFLPE varied across different levels of individual student ability; and (c) elucidate whether the BFLPE could be moderated by particular individual student characteristics. Results confirmed the external validity and universal applicability of the BFLPE by demonstrating that it was not only evident in developed countries and individualist countries, but also in economically developing countries and collectivist countries. The BFLPE was shown to vary across different ability levels, with larger BFLPEs being associated with students who were higher in ability. Some individual student characteristics showed evidence of a moderating effect on the BFLPE. Although many of these effects were trivial, especially given the power engendered by the large sample used in the current study, constructs such as high levels of anxiety, memorisation, and a preference for a cooperative learning environment were shown to be associated with larger BFLPEs. These results are discussed in more detail in the sections that follow.

The External Validity and Generalizability of the BFLPE

Compared to those that are culturally dependent, theories that span cultures are regarded as more externally valid, more universal, and more generally applicable. As Schwartz and Bilsky (1990) noted, “Theories that aspire to universality…. must be tested in numerous, culturally diverse samples” (p. 878). In order to demonstrate the universality and external validity of the BFLPE, the current study applied Schwartz
and Bilsky’s conditions by analysing data from 41 countries, including economically developing and developed countries and individualist and collectivist countries.

Previous research has demonstrated that the BFLPE is evident mainly in developed countries and individualist countries (e.g., Craven et al., 2000; Marsh & Craven, 2002; Marsh et al., 2001; Marsh & Rowe, 1996; Marsh et al., in press; Mulkey et al., 2005). Some studies have demonstrated the existence of the BFLPE in developing countries and collectivist countries, but these have been few. Marsh and Hau’s (2003) study was the most impressive of these studies, finding evidence of a BFLPE in 26 countries, and is the study on which the current investigation was based. However, the current study greatly extended Marsh and Hau’s work by: (a) expanding the number of countries tested; (b) including many more collectivist and developing countries; (c) assessing the moderating effect of cultural orientation and stage of economic development on the BFLPE; and (d) explicating moderators of the BFLPE.

In the current study, the BFLPE was evident not only in individualist and economically developed countries, but also in economically developing and collectivist nations. This was demonstrated in both the cross-national sample and on a country-by-country basis. The cross-national sample of 41 countries was comprised of individualist countries, economically developed countries, economically developing countries, and collectivist countries. Analyses using this sample demonstrated that individual ability had a positive association with math self-concept and that, controlling for individual ability, school-average math ability had a negative association with math self-concept. This negative association is characteristic of the BFLPE. Thus, in the cross-national sample, students in high-ability schools had lower math self-concepts than their equally able counterparts in low-ability schools. These findings are consistent with previous research in developed and individualist countries (e.g., Craven et al., 2000; Marsh & Craven, 2002; Marsh et al., 2001; Marsh & Rowe, 1996; Marsh et al., in press; Mulkey et al., 2005), collectivist countries (Marsh et al., 2000; Zeidner & Schleyer, 1998), and developing countries (Marsh & Hau, 2003).

In the cross-national sample, consistent with the findings of Marsh and Hau (2003), there was a significant but small variation in the size of the BFLPE from
country to country. To ascertain whether a BFLPE was displayed in the individual
countries that formed the cross-national sample, the BFLPE was evaluated on an
individual country basis. To accomplish this, separate multilevel regressions were
conducted for each country. In all 41 countries examined individually, controlling for
individual ability, school-average math ability had a negative relation with math self-
concept (the BFLPE), and this effect was statistically significant in 38 of these
countries. As such, these results replicate those of Marsh and Hau (2003), but also
extended them considerably by including many more developing countries and
collectivist countries. Of the three countries displaying non-significant effects, one
could be regarded as individualist (Ireland) and one as collectivist (Korea). The third,
Iceland, did not receive a score on the Individualism Index, used to classify
countries. All three countries are regarded as economically developed (see World
Bank, 2007). Consequently, it appears that the BFLPE is not restricted to developed
countries or to individualist countries but also occurs in collectivist countries and
developing nations.

To further strengthen claims concerning the external validity of the BFLPE,
analyses were conducted to ascertain whether cultural orientation (individualism/
collectivism) or status of economic development (developed/developing) could
modify the BFLPE. Results indicated that the BFLPE was present irrespective of
economic development or of a country’s cultural orientation.

The current study tested the external validity and generalisability of the BFLPE
cross-nationally in 41 countries, on an individual country basis, and by assessing the
moderating effects of cultural orientation and status of economic development.
Results of this large-scale comprehensive test of the BFLPE indicate that the BFLPE
is not only a symptom of developed countries or individualist societies, but it is also
evident in developing nations and in the more collectivist countries of the world.

**Individual Ability and the BFLPE**

As with previous BFLPE research (e.g., Marsh, 1991; Marsh & Parker, 1984;
Marsh et al., in press), in all instances individual ability positively predicted math
self-concept, whereby students of higher ability had higher math self-concepts.
Additionally, consistent with findings by Marsh and Hau (2003), there was evidence of a quadratic effect for individual ability. Although math self-concept increased as ability increased, the findings were suggestive of a gradual increase for students of below-average ability and a stronger increase for students of above-average ability.

Previous research assessing the moderating influence of individual ability on the BFLPE has been inconsistent in direction and in statistical significance (Coleman & Fults, 1985; Marsh et al., 1995; Marsh & Hau, 2003; Marsh & Rowe, 1996; Reuman, 1989). In the present investigation across the span of ability levels, students who attended high-ability schools had lower math self-concepts than equally able students in low-ability schools. This decline in math self-concept was slightly worse, however, for high-ability students in high-ability schools. Slightly larger BFLPEs were associated with students in high-ability schools whose ability levels were one to two standard deviations above the mean.

On the basis of social comparison theory, one could predict that high-ability environments would not perturb high-ability students. In these environments, high-ability students would have low-ability students with whom to compare themselves and presumably these downward comparisons could bolster their self-evaluations (see Chapter 3). This was not the case for the students in the current study. Marsh et al. (1995) argued that high-ability students might not be able to excel in all of their subjects and this may result in them being dissatisfied with their performance. This may have been the case in the current study resulting in high-ability students suffering more from the BFLPE than low-ability students.

Notwithstanding the significant moderating effect of individual ability, the linear ability interaction effect was small, and the quadratic ability interaction effect was even smaller, so small in fact that it was not visually evident in a graph of the interaction. Given the size of the sample and the power of the tests associated with a large sample, such small effects mean that implications arising from these results can only be suggestive and lead to the conclusion that the BFLPE generalises well over different ability levels.
Individual Student Characteristics and the BFLPE: SES, Individual Differences in Learning, and Perceptions of the Learning Environment

In the current study, particular individual student characteristics were assessed to determine if any alleviated or exacerbated the negative effects of the BFLPE. These potential moderators included SES, individual differences in learning, and individual perceptions of the learning environment, and are discussed in the following sections.

SES

Previous research has linked higher SES levels with more positive educational outcomes and higher academic self-concept (e.g., DeGarmo et al., 1999; Marks et al., 2006; Organisation for Economic Cooperation and Development, 2001a; Pong & Ju, 2000). Results from the current study indicated that students from higher SES backgrounds had higher math self-concepts. These findings are consistent with those of Marsh and Parker (1984) and Bachman and O’Malley (1986) who found that SES had a significantly positive effect on academic self-concept. However, the association between SES and math self-concept in the current study was weak, suggesting that the contribution of SES to the prediction of math self-concept is minimal.

When SES variables were tested as moderators of the BFLPE, only two SES indicators produced significant results: highest parental occupation and home educational resources. The pattern of results suggested that slightly smaller BFLPEs were associated with students in high-ability schools when a parent was employed in a more prestigious occupation or when there were more educational resources in the home. Perhaps having a parent in a high-level occupation and having more educational resources in the home inspire students academically so as to moderate the negative effects of the BFLPE. However, as was the case for individual ability, although statistically significant, these effects were small in size. As the remaining SES interactions were not statistically significant, these results are merely suggestive and should be considered in the context of the sample size and the power of the tests. Accordingly, these analyses provide evidence that the BFLPE generalises well over different levels of SES.
Individual Differences in Learning

With the exception of math anxiety, all individual differences in learning variables had a positive association with math self-concept. Students had higher math self-concepts if they used learning strategies, such as elaboration, memorisation, or control strategies. They also had higher math self-concepts if they were extrinsically or intrinsically motivated, if they believed in their own self-efficacy, or if they espoused a preference for a competitive or cooperative learning environment. That these individual differences in learning variables were positively related to math self-concept is consistent with research that has demonstrated that they are associated with beneficial educational outcomes (Boekaerts, 1997; Boekaerts & Cascallar, 2006; Camahalan 2006; Ginsburg & Bronstein, 1993; Gottfried, 1985, 1990; Marsh, Hau, et al., 2006; Multon et al., 1991; Organisation for Economic Cooperation and Development, 2001a; Otis et al., 2005; Ryan & Deci, 2000). Conversely, students who were highly anxious about their math studies had lower math self-concepts, a finding consistent with research demonstrating a negative relation between math anxiety and math ability perception, a construct similar to math self-concept (Meece et al., 1990).

When individual differences in learning variables were examined as moderators of the BFLPE, significant negative associations were found for math anxiety, a preference for a cooperative learning environment, and the use of memorisation as a learning strategy. Students were slightly more likely to be victims of the BFLPE if they were highly anxious about their math studies, if they preferred to learn in a cooperative environment, or if they used memorisation as a learning technique.

The pattern of results for math anxiety indicated that students who exhibited low math anxiety had similar math self-concepts in low- and high-ability schools. Thus low anxiety students were not affected by the BFLPE. Students in high-ability schools with average anxiety levels showed a slight decline in their math self-concepts, but highly anxious students in high-ability schools had considerably lower math self-concepts than highly anxious students in low-ability schools. Thus, a larger BFLPE was associated with highly anxious students compared to those who exhibited average or low anxiety levels. This finding has intuitive appeal. Students who are highly anxious about their academic performance may not achieve to their
potential. If they are not performing to their best, anxious students may feel negative about their abilities, and this may be compounded if they attend a high-ability school where they are surrounded by high-achieving peers. The result may be a lower academic self-concept – the BFLPE.

The BFLPE was greater for students who indicated a preference for working in a cooperative environment, compared to those who espoused a cooperative learning environment to a lesser extent. Students who prefer cooperative environments enjoy working and sharing ideas with others. However, high-ability schools often foster a competitive spirit and this may be especially distressing for students who prefer to work cooperatively. Cooperatively oriented students may feel vulnerable in competitive environments, making them feel less capable and resulting in a drop in their self-concepts.

Similarly, the BFLPE was larger for students who used the learning strategy memorisation to a greater extent. These students use rehearsal techniques to learn material, and so tend to learn by rote rather than by gaining a deep understanding of the concepts. Perhaps students in high-ability schools who learn by rote may not feel as confident about their abilities. In high-ability schools, school work may need to be addressed that requires the application of concepts for which understanding is necessary and which cannot be solved using rote learning. In such circumstances students who use memorisation techniques may often find themselves out of their depth, and this may be reflected in lower self-concepts.

Significant positive interactions were demonstrated for elaboration, extrinsic and intrinsic motivation, and self-efficacy. The BFLPE was slightly smaller if students used elaboration methods to learn, if they were extrinsically or intrinsically motivated, or if they felt a degree of self-efficacy as regards their math studies. Students who used elaboration as a learning technique would have thought deeply about math concepts by relating them to other material. This may have made them more confident about their abilities, thus buffering them somewhat against the BFLPE. Similarly, being more motivated to study math and feeling that one can succeed in math may have inspired confidence in students that helped to counteract the BFLPE.
Although all of these interaction effects were statistically significant, many were small. When interactions for elaboration, extrinsic and intrinsic motivation, and self-efficacy were graphically represented, it became clear that the significant interactions were only produced by the power engendered from the large sample size, as the interaction effects were barely noticeable. This suggests that the BFLPE is reasonably consistent across these specific student characteristics. The interaction effects for memorisation and for cooperative learning were larger and hence of more importance. However, they, too, should be considered in the context of the large sample size, and any implications drawn from them treated cautiously. The math anxiety interaction was a reasonably large effect and so may provide information for alleviating the negative effects of the BFLPE.

**Individual Perceptions of the Learning Environment**

Previous research has demonstrated that more positive perceptions of the learning environment are related to better educational and mental health outcomes (Crosnoe, et al., 2004; Ozer, 2005). Consistent with these findings, results of the current study indicated that students had higher math self-concepts if they felt a sense of belonging to their school, if they held more positive attitudes to school, or if they had good relationships with their teachers.

When individual perceptions of the learning environment variables were examined as moderators of the BFLPE, two variables produced significant, though small, effects. Smaller BFLPEs were associated with students feeling a sense of belonging to the school, or with students holding positive attitudes towards school. Students in high-ability schools may have felt more connected to their schools because they did not feel different to other students academically, and this feeling of connection may have reduced the BFLPE somewhat. Students in high-ability schools who held more positive attitudes to school may have regarded attending such schools as being beneficial to them, with the result that the BFLPE was moderated for them. Nevertheless, the size of these interaction effects was trivial. When examined graphically, whereas the attitudes towards school interaction was noticeable, the sense of belonging interaction was barely discernable. As previously mentioned, these results and their attendant implications have to be considered in conjunction with the power associated with a large sample. Thus, it appears that the BFLPE is
reasonably consistent across the individual perceptions of the learning environment as considered in the current study.

**Section Summary**

Significant interaction effects were found for many of the potential moderators of the BFLPE. Higher parental occupation and more home educational resources were two SES variables associated with smaller BFLPEs. Smaller BFLPEs were also associated with students who: (a) used elaboration techniques; (b) were motivated either extrinsically or intrinsically; (c) felt a sense of self-efficacy in math; (d) had more positive attitudes to school; or (e) felt a connection to the school. Nevertheless, these interaction effects were trivial. Larger effects were found for students who used memorisation strategies, or who preferred cooperative learning environments. In both these cases, students were more affected by the BFLPE. However, in a more moderately sized sample none of these effects may have been statistically significant. The current sample comprised over a quarter of a million students. Large samples such as this provide much power and can produce small significant effects that would not otherwise reach significance (Howell, 1997). The significant interactions for the moderating factors found in the current study were generally small and probably reached significance due to the large sample size. This being the case, although these moderators can be used to provide suggestions as to how the BFLPE may be overcome, they should be treated cautiously. On the contrary, the size of the effect for math anxiety was larger and thus may provide information on overcoming the BFLPE.

**Strengths and Limitations**

A major strength of the current study is that it encompassed a very large, culturally diverse sample. This sample was ideal for assessing the external validity of the BFLPE as it included numerous countries incorporating different cultural orientations and countries at different stages of economic development. Using such a sample contributed to providing strong evidence for the universality of the BFLPE.

Another key strength of the current study is that it extended Marsh and Hau’s (2003) work in valuable ways. Firstly, Marsh and Hau’s sample comprised 26
countries. The current study expanded on this sample size by increasing the number of countries investigated to 41. Marsh and Hau included only five collectivist countries and six economically developing countries. The sample on which the current study was based included 16 countries that could be considered as collectivist and 14 as developing, thus greatly increasing the number of both collectivist and developing countries examined. The current study also tested whether the BFLPE was moderated by culture or economic development, an issue not covered by Marsh and Hau. Additionally, in the original Marsh and Hau study, academic self-concept was the outcome variable. The current study employed math self-concept as the outcome variable, thus underscoring the extreme domain specificity of the BFLPE. Lastly, the current study also extended Marsh and Hau’s work by explicating an extensive range of BFLPE moderators. In these ways, the current study not only replicated Marsh and Hau’s findings but also extended them in significant and important ways.

Additionally, the current study employed multilevel modelling to analyse these data. When samples are comprised of groups nested within other groups and when these groups have not been randomly assigned, independence of observations cannot be assumed. Especially in the case of students within schools as in the current study, it is likely that students in one school were more similar to each other than they were to members of another school. By using multilevel modelling, the variance from each group level could be partitioned, resulting in more accurate statistical information, thus overcoming any lack of independence in the data.

A limitation of the current study is that it was based on correlational data, so no causality can be inferred. Educational studies such as this routinely rely on correlational data, as it would be not only impossible, but also unethical to randomly assign students to different schools. However, the causality issue is an important one. Results of the current study assume that individual and school-average ability precede self-concept, but research has shown that academic achievement and self-concept are reciprocally related (see Chapter 2). So, it is unknown whether students entered high-ability schools with lower self-concepts, or whether their self-concepts were lowered by virtue of attending such schools. For the causal relation between self-concept and attending high-ability schools to be explicated fully, self-concept
should be measured before, during and after attending a high-ability school. Especially in the case of the moderator variables, such as anxiety, it is unknown from the current study whether students entered high-ability schools with existing high levels of anxiety or whether anxiety was increased by attending a high-ability school. As a result, interpretations of the effects of the moderator variables, such as anxiety, are rendered problematic. This limitation could be overcome by longitudinal causal modelling studies in which students are assessed on these variables before they attend schools that differ substantially in terms of school-average ability. Countries such as Germany and Hong Kong would be ideally suited to this type of research. In these countries students are educated in reasonably homogeneous primary schools and continue their education in highly academically segregated high schools. Students could be assessed in their last year of primary school and then during the course of their high school education. In this way causality could be determined.

**Implications for Theory, Practice, and Research**

By demonstrating that the BFLPE extends to economically developing countries and collectivist countries, the current study not only extends previous research, which has focussed primarily on developed countries and individualist nations, but also has important theoretical implications for the BFLPE. By including such a large eclectic mix of countries, the present investigation has encompassed both of Schwartz and Bilsky’s (1990) conditions regarding the universality of a theory, by showing the BFLPE to be evident in numerous culturally diverse countries. In doing so, the current study has demonstrated that the BFLPE is externally valid and generally applicable and has placed the BFLPE as a universal theory, relevant to people everywhere.

That the BFLPE is a universally applicable theory has implications for educational policy makers around the world. In many countries worldwide (e.g., Hong Kong, Germany, Australia) high-ability students are increasingly being taught in academically selective schools. Education policies globally emphasise the need to ensure that high-ability students reach their full potential (e.g., Department for Education and Skills UK, 2005). The current results, coupled with previous research
on the BFLPE and the REM (see Chapter 2), suggest that segregation on the basis of academic ability, far from assisting students to reach their full potential, may actually hinder that process. Research supporting the REM (Marsh & Craven, 2005, 2006; Marsh & Yeung, 1997a; Valentine & DuBois, 2005) has demonstrated that academic self-concept and achievement are reciprocally related: Increases in one are associated with increases in the other and vice versa. This would suggest that for high-ability students to develop to their full potential, academic self-concept and achievement must be improved simultaneously. However, as our results show, when students are educated in high-ability schools their academic self-concepts suffer, and taking the REM to its logical conclusion, so must their performance. Hence, academic segregation may not be the optimal environment for high-ability students to realise their full potential. If, as these results imply, segregation on the basis of academic ability is detrimental to a student’s self-worth, irrespective of culture and ability level, then the task of limiting the negative effects of the BFLPE is a global challenge – one that it is imperative that researchers, teachers, and policy makers take up worldwide.

Results for the potential BFLPE moderators have implications for practice, but these should be considered in light of earlier comments regarding causality and the size of the interaction effects, given the large sample size. Notwithstanding these caveats, the finding that the BFLPE was minimal for students low in anxiety has major implications for practice and future research. Discovering whether students in high-ability schools have pre-existing anxiety problems or whether attending high-ability schools causes students to be anxious should be a top priority for future research. Additionally, a major focus should be placed on interventions aimed at aiding students in high-ability schools to overcome their academic anxieties. High-ability schools, too, should be conscious that anxiety aggravates the BFLPE and endeavour to create environments where anxiety is reduced. Especially when one considers that higher levels of anxiety have been linked with lower achievement (e.g., Chapell et al., 2005; Martin, 2003; Zeidner & Schleyer, 1998), these matters should be given priority.

Other moderators, too, have implications for practice. Perhaps teaching students how to use elaboration instead of memorisation as a learning strategy may help to
reduce the BFLPE. Interventions aimed at students who prefer to work cooperatively may allow them to cope with the competitive environment of a high-ability school. Alternatively, high-ability schools could endeavour to promote more cooperative learning environments. Additionally, the BFLPE may also be counteracted by strategies that: (a) increase motivation; (b) encourage students to believe that they can achieve; (c) promote positive attitudes to school; and (d) instil a sense of connection to the school. However, given that self-concept is central to the BFLPE, strategies that focus on enhancing students’ self-concepts in specific subject domains and that serve to protect rather than undermine self-concept offer the most potent practical potential for counteracting the BFLPE.

Additional areas for future research include studies to test the effectiveness of different learning strategies in high-ability schools and the causal relation between the two. Apart from reducing academic self-concept, attending a high-ability school may also have a negative impact on how students learn. In the context of the current study, attending such a school may impact on the individual differences in learning factors. It may be that attending a high-ability school undermines these potentially positive learning strategies, rendering them less effective. As the current sample contained data from only a single time period, it was not possible to test the causal ordering of such variables. A longitudinal design is therefore necessary to test any causal ordering that may exist.

Additionally, other individual differences could be examined in future research. For example, a more detailed study of individual student motivation, including mastery and performance achievement goals, may prove fruitful. Personality differences may be another avenue for future research to pursue. Perhaps more extroverted students or those who are more emotionally stable may suffer less from the negative effects of the BFLPE. At the school level, not being urged by teachers (or parents) to achieve academically may prove to lessen the BFLPE. Further, due to constraints of the database, the current investigation could not elucidate the psychological processes that cause the BFLPE. Future research should examine these processes and develop intervention strategies to counter the negative effects of the BFLPE.
Summary

The current chapter described Study 1 of the present investigation. The background to Study 1 was presented and the aims, hypotheses, and research questions for Study 1 were outlined. Based on previous research and theory, described in Chapter 2, a brief rationale was given for each of the hypotheses and research questions to be tested in the current study. The chapter continued by outlining the methodology used to conduct Study 1, including descriptions of the participants, measures, and the procedure used. A detailed description of the statistical analyses used in all three studies of the present investigation was included. Results of analyses were provided and these indicated that: (a) the BFLPE was externally valid and universally applicable; (b) the BFLPE varied across ability levels, with larger BFLPEs being associated with students of higher ability; and (c) the BFLPE was moderated by various constructs. In particular, the BFLPE was reduced for students who used elaboration, who were motivated, who had a sense of self-efficacy, who had positive attitudes to school, and who felt connected to their school. The BFLPE was heightened for students who were anxious, who used memorisation, and who preferred a cooperative learning environment. The chapter concluded by presenting a discussion of these results. The strengths and limitations of Study 1 were discussed and implications of the findings for theory, practice, and future research were also offered.
CHAPTER 5

STUDY 2: DO THE POSITIVE EFFECTS OF UPWARD COMPARISON CO-EXIST WITH THE BFLPE?

Introduction

The purpose of this chapter is to present Study 2 of the present investigation, which aims to extend current knowledge of the BFLPE by investigating a contradiction that exists between findings from social comparison studies and those of the BFLPE. As described in Chapter 3, the BFLPE has been explained in social comparison terms. In essence, it has been suggested that it is the forced comparison with classmates who are performing better that leads to a lowered academic self-concept. Conversely, social comparison research has found that students nominated a social comparison target who slightly outperformed them in class with a beneficial effect on course grades, but with no effects on self-evaluation (Blanton et al., 1999; Huguet et al., 2001); see Chapter 3. This contradiction between BFLPE results in regard to academic self-concept and social comparison findings in regard to performance is the focus of the current study, as to date few studies have evaluated the BFLPE and the effects of upward social comparison in educational settings together in the same study. Hence, this component of the current research sought to bring these two strands of research together. The overarching purpose of Study 2 was to further analyse two social comparison studies (Blanton et al., 1999; Huguet et al., 2001) that demonstrated that upward comparisons improved performance, in order to identify whether the negative effects of the BFLPE on self-concept could co-exist with, or be moderated by, the positive effects of upward comparison.

The current study began by replicating the results of Blanton et al. (1999) and Huguet et al. (2001), using both multiple regression and multilevel modelling. It continued by exploring whether these results were also evident in new data, collected by Huguet et al. (2001) at the same time as that used in their initial study, but not previously reported. These new data provided a significant increase in the number of participating students, classes, and schools, thereby allowing for greater generalisability of the results and the application of more appropriate statistical
analyses (i.e., multilevel modelling). The study concluded by investigating whether the BFLPE was evident in the social comparison data of Blanton et al. and Huguet et al.’s expanded data in an attempt to explicate whether the BFLPE was moderated by selected comparisons or whether the two effects co-existed.

The current chapter aims to present for Study 2: (a) the nature of the problem to be investigated, (b) the background to the problem, (c) the overarching aims, (d) specific hypotheses to be tested, (e) research questions to be addressed, (f) the rationale for hypotheses and research questions, (g) the methods employed, and (h) results of the investigation in the context of the hypotheses and research questions posed. The chapter concludes with a discussion of the results and the implications thereof for theory, research, and practice. As Study 2 was a further analysis of two previously published studies, for clarity analyses for the two studies are presented separately. Study 2a is the further analysis of the Blanton et al. (1999) data and Study 2b is the further analysis of the Huguet et al. (2001) data. Thus, for each study, aims, hypotheses, research questions, rationales, methodology, and results are presented separately. Results for both studies are considered together in the discussion section.

**Statement of the Problem**

Does the BFLPE exist in the same data that demonstrated the positive effects of selected upward comparisons on performance? Does the BFLPE co-exist with these findings, or is it moderated by selected upward comparisons? Does the BFLPE vary as a function of individual ability?

**Background**

Wheeler and Suls (2005) were the first to emphasise an apparent discrepancy between the BFLPE results of Marsh and his colleagues (e.g., Marsh, 2005; Marsh & Craven, 2002; Marsh & Hau, 2003; Marsh & Parker, 1984; see Chapter 2) in relation to social comparison and academic self-concept, and the social comparison findings of Blanton et al. (1999) and Huguet et al. (2001) in relation to actual performance. Whereas the latter social comparison studies (described in Chapter 3) found that
upward social comparisons with a high-achieving other had beneficial effects on subsequent performance, the BFLPE model postulates that social comparisons are responsible for lowered academic self-concepts, and in turn, lower academic self-concepts have been associated with poorer academic performance (e.g., Marsh & Yeung, 1997a).

Accordingly, Wheeler and Suls challenged Marsh to reconcile these two seemingly divergent results (J. Suls, personal communication, September 11, 2003). In response, Marsh (H. Marsh, personal communication, 12th November, 2003) noted that because the Blanton et al. (1999) and Huguet et al. (2001) studies did not specifically evaluate academic self-concept and, in particular, did not include a measure of class- or school-average achievement that is central to tests of the BFLPE, neither study provided a test of the BFLPE nor explained how it related to the processes which were evaluated in these two studies. Marsh continued by noting that differences in class-average school grades had been scaled away in the Blanton et al. (1999) and Huguet et al. (2001) studies either by standardising grades or by centring the effects separately within each class. Thus, he suggested that a BFLPE would be present for self-evaluations if an appropriate measure of class-average achievement were available. (It should of course be noted that the purpose of both these studies was not to test for the BFLPE, but rather to examine the effect of comparisons on performance, so it is not surprising that these conditions were not met). At the same time, Marsh invited the authors of Blanton et al. (1999) to collaborate with him in a further analysis of their findings in order to explore whether the BFLPE co-existed with their original results. Subsequently, Huguet et al. (personal communication, April 29, 2004), in response to Wheeler and Suls’ identification of the apparent discrepancy in the literature and Marsh’s proposed further analysis, suggested that all parties engage in a joint venture to investigate the relation between the BFLPE and social comparison choices.

Consequently, Study 2 represents the collaborative effort undertaken to reconcile the contradiction between BFLPE results and those of social comparison research. It investigated whether lowered self-evaluations of academic ability, resulting from being in a higher ability class, can co-exist with, or be moderated by, upward comparisons that produce enhanced performance. Study 2 is one of the first studies
to evaluate the BFLPE simultaneously with the effects of social comparison in educational settings. Thus, Study 2 represents one of the first steps in exploring the theory that social comparison processes underlie the BFLPE, by focusing specifically on the issue of selected individual comparison targets versus forced generalised comparison targets in an attempt to explicate their effects on self-evaluations of competency.

In summary, Study 2 aimed to determine whether social comparisons could moderate the BFLPE, or whether the negative effects of the BFLPE co-existed with the beneficial effects of upward social comparisons. In this manner, these two theoretical perspectives were brought together in the context of a single study, whereby the original Blanton et al. (1999) and Huguet et al. (2001) data were analysed from a BFLPE perspective in an attempt to resolve the apparent contradiction in the literature.

**Overarching Aims**

The principal overarching aim of Study 2 was to ascertain whether selected individual comparisons could moderate the BFLPE, or whether the negative effects of the BFLPE co-existed with the beneficial effects of upward comparisons on performance. An ancillary aim was to determine whether the BFLPE differed depending on individual ability levels. In doing so, Study 2 further analysed the Blanton et al. (1999) data, the Huguet et al. (2001) data, and analysed an expanded data set on which the original Huguet et al. data were based. Study 2a of the present investigation presents a reanalysis of the Blanton et al. data, and Study 2b presents a reanalysis of the Huguet et al. data including the expanded data set. Specific aims for Studies 2a and 2b are presented within the sections devoted to each of these studies that follow.
Study 2a: Further Analysis of Blanton et al. (1999)

Aims

Study 2a aimed to further analyse the Blanton et al. (1999) data to ascertain whether individual selected comparisons that have been shown to enhance performance moderate the BFLPE, or co-exist with the BFLPE. More specifically, Study 2a aimed to:

1. Analyse Blanton et al.’s results for performance and self-evaluation based upon multiple regression in order to replicate their original results;
2. Utilise multilevel modelling to analyse Blanton et al.’s original results for performance and self-evaluation in order to demonstrate that their original results could be replicated using this stronger statistical technique;
3. Conduct further analyses of Blanton et al.’s data utilising a multilevel modelling approach to test whether lower academic self-evaluations were present as a result of being in a high-ability class (the BFLPE);
4. Utilise the Blanton et al. data to compare and contrast the BFLPE and predictions from social comparison theory in order to ascertain whether the BFLPE is moderated by upward comparisons or whether both theoretical propositions co-exist; and
5. Explore the moderating effect of individual student ability on the BFLPE in the Blanton et al. data in order to explicate whether the BFLPE varies as a function of individual ability.

Statement of the Hypotheses and Research Questions

Previous theory and research were used to generate hypotheses, but when this was not possible research questions were devised. The numbers relating to hypotheses and research questions are related to the aim from which they are derived. Hence, for example, Hypothesis 1.2 is the second hypothesis based on the first aim, and Research Question 4.1 is the first research question based on the fourth aim. The hypotheses and research questions that follow were formulated to achieve the aims of Study 2a.
Hypothesis 1.1: Replication of Blanton et al.’s (1999) Results Utilising Multiple Regression – Grades

The results of the Blanton et al. (1999) study will be replicated whereby, based on multiple regression, for all seven school subjects (biology, Dutch, English, French, geography, history, math) T2 grade will be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice.

Hypothesis 1.2: Replication of Blanton et al.’s (1999) Results Utilising Multiple Regression – Self-Evaluation

The results of the Blanton et al. (1999) study will be replicated whereby, based on multiple regression, for all seven school subjects (biology, Dutch, English, French, geography, history, math) self-evaluation will be significantly positively predicted by T2 grade but T2 comparison choice will not be a significant predictor of self-evaluation for any of the academic subjects.

Hypothesis 2.1: Replication of Blanton et al.’s (1999) Results Utilising Multilevel Modelling – Grades

When using multilevel modelling Blanton et al.’s (1999) results will be replicated for all seven school subjects, whereby T2 grade will be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice.


Using multilevel modelling Blanton et al.’s (1999) results will be replicated for all seven school subjects, whereby self-evaluation will be significantly positively predicted by T2 grade but T2 comparison choice will not be a significant predictor of self-evaluation.


When the Blanton et al. (1999) data are analysed using multilevel modelling, it is predicted that for each school subject, self-evaluation as measured in the Blanton et
al. (1999) study will be significantly positively predicted by individual ability as measured by T1 grade. This relation is depicted in Figure 5.1.


It is predicted that when the Blanton et al. (1999) data are analysed using multilevel modelling, for each school subject, self-evaluation as measured in the Blanton et al. (1999) study will be significantly negatively predicted by class-average ability, as measured by T2 class-average grade – the BFLPE (see Figure 5.1).


In the Blanton et al. (1999) data, for each school subject, it is expected that the main effect of the comparison person’s grade, as measured by his or her T2 grade, will not have a statistically significant relation with self-evaluation as measured in the Blanton et al. (1999) study (see Figure 5.1).

![Figure 5.1. Relation of Individual Ability, Class-Average Ability, and Comparison Choice with Self-Evaluation](image-url)

Does comparison choice moderate the BFLPE? More specifically, for each school subject, does the comparison person’s grade, as measured by his or her T2 grade, moderate the effect of class-average ability, as measured by T2 class-average grade, on self-evaluation as measured by Blanton et al. (1999)? If there is no moderation, then do the positive effects of comparison choice on performance and the negative effects of class-average ability on self-evaluation co-exist?

Research Question 5.1: Individual Ability as a Moderator of the BFLPE in Blanton et al.’s (1999) Data Based on Multilevel Modelling

Is the BFLPE moderated by individual ability level? For each school subject in the Blanton et al. (1999) data, what is the relation between individual ability, as measured by T1 grade, and class-average ability, as measured by T2 class-average grade, on self-evaluation as measured in the Blanton et al. (1999) study?

Rationale for Hypotheses and Research Questions

Rationale for Hypotheses 1.1 and 1.2: Replication of Blanton et al. (1999) Utilising Multiple Regression – Grades and Self-Evaluation

As these components of the current investigation are using the same statistical technique as those of the original studies, the same pattern of results is anticipated for both grades and self-evaluation.


Multilevel modelling takes into account the nested nature of data when calculating coefficients and standard errors. Hence, traditional single level techniques that ignore a multilevel structure are likely to be invalid as they not only violate assumptions of independence, but they also increase the likelihood of finding statistical significance where none exists (Hox, 2002); see Chapter 4. Although the original results for this study were based on single level multiple regressions, they were generally robust. Thus, it is anticipated that similar results will be demonstrated
for grades and self-evaluation when multilevel regression analyses are used instead of single level regressions.


The BFLPE theoretical model proposes that individual ability is positively related to academic self-concept, and this has been demonstrated empirically (e.g., Craven et al., 2000; Marsh, 2005; Marsh & Craven, 2002; Marsh & Hau, 2003; Marsh et al., 2001; Marsh et al., in press; Mulkey et al., 2005). While academic self-concept per se was not measured in this study (i.e., not the usual measure of academic self-concept used in BFLPE studies), the self-evaluation measure used was considered an approximate measure of academic self-concept (see Methodology section below). Thus, it seemed appropriate to conclude that the same relation that exists between individual ability and academic self-concept would also be found between individual ability and the single measure of self-evaluation used in the current study.


BFLPE research has consistently found that the relation between class-average ability and academic self-concept (self-evaluation in this study) is significantly negative (e.g., Craven et al., 2000; Marsh, 2005; Marsh & Craven, 2002; Marsh & Hau, 2003; Marsh et al., 2001; Marsh et al., in press; Mulkey et al., 2005). Although as noted above, in the absence of an academic self-concept measure a single measure of self-evaluation was used in the current study, it is predicted that a similar pattern of results will be demonstrated in the present investigation.


Comparing one’s achievements with others can either be inspiring (Lockwood & Kunda, 1997) or deflating (e.g., Collins 1996, 2000; Craven et al., 2000; Lockwood & Kunda, 1997; Marsh, 2005; Suls & Wheeler, 2000). However, in the Blanton et al.
(1999) study the comparison person’s grade had no effect on self-evaluation for any of the academic subjects tested. As the current study is a further analysis of the Blanton data, it is expected that results will be similar to that study and hence the comparison person’s grade will not be statistically significantly related to self-evaluation.


As noted in the rationale for Hypothesis 4.1, research examining the effect of comparisons on self-evaluations can either be inspiring or deflating. However, no previous research that the author is aware of has examined the relation between the comparison choice-by-class-average ability interaction and self-evaluation. Whereas the social comparison studies of Blanton et al. (1999) and Huguet et al. (2001) demonstrated that chosen upward comparisons improved academic performance, the BFLPE model predicts that forced comparisons with the generalised other lead to lowered self-evaluations of ability (see Chapters 2 & 3). This begs the question of whether chosen comparisons with a specific other person can moderate the negative effects of the BFLPE. Will such a selected comparison reduce or increase the BFLPE? Alternatively, comparison choice and class-average ability may not interact, in which case the two effects (the positive effects of upward comparison on performance, and the BFLPE) may be said to co-exist. Hence, given these issues, and as past research is unable to provide guidance for formulating clear predictions, a research question is posed to illuminate these questions.

**Rationale for Research Question 5.1: Individual Ability as a Moderator of the BFLPE in Blanton et al.’s (1999) Data Based on Multilevel Modelling**

In Chapter 2 it was noted that the evidence for the moderating effect of individual ability on the BFLPE has been conflicting. Some studies have shown little or no evidence of a BFLPE for high-ability students in academically selective classes (e.g., Coleman & Fults, 1985); others have noted that high-ability students in homogeneous high-ability classes felt the negative effects of the BFLPE more than low-ability students in low-ability classes (Reuman, 1989); while still others have shown that students of all ability levels suffer the negative effects of the BFLPE
When individual ability was examined as a moderator of the BFLPE in Study 1, larger BFLPEs were associated with students of higher ability. However, due to the small size of the interaction and the power engendered by the large sample in that study, it was concluded that the BFLPE generalised reasonably well over different ability levels. Given the conflicting results surrounding individual ability as a moderator of the BFLPE, no a priori predictions are made regarding this interaction; instead, a research question is posed to clarify this matter.

Methodology

Participants

In the Netherlands, students begin their first year of high school when they are about 12 years old. Dutch schools are highly segregated on the basis of individual ability (see Chapter 2). At the time of the Blanton et al. (1999) study, the Dutch educational system had five different tracks, although the schools in the Blanton et al. study only offered the three highest tracks. Depending on the track students are enrolled in, they attend high school for between four and six years. At least for the first year, students attend all courses with the same group of classmates. Participants in the Blanton et al. (1999) study were 876 students (432 boys and 444 girls) in the first year of high school, across four Dutch schools. As dates of birth were not accessed, the original researchers estimated that the median age of the sample was between 12 and 13 years.

In each high school, all students in their first year were given the opportunity to participate. None refused, but those who were absent when the primary questionnaire was administered, or who no longer attended the school, were not included in the sample. Students were not randomly assigned to either the school or class, and may have chosen their school for a number of reasons, ranging from the ability level of the school to the proximity of the school to the student’s home. The schools in this study would typically be considered to be of medium to high-ability levels (although within the schools individual classes were rated as low, medium, and high-ability; see section on modified grades below). Schools were selected to participate on the
basis of their interest in contributing to the research and in receiving a financial reward. Classes consisted of between 23 to 30 students, and there were between 7 to 11 classes in each school (see Blanton et al., 1999, for full details).

**Measures**

*Grades.* Grades, awarded by class teachers on a 10-point response scale with 10 being the highest grade, were used to determine performance. Descriptors related to these grades were “very bad for a 3, sufficient for a 6, and good for an 8” (Blanton et al., 1999, p. 423). Grades were accessed for seven academic subjects: biology, Dutch, English, French, geography, history, and math. Average scores for these subjects, as reported by Blanton et al., are reproduced in Table 5.1. These grades were subsequently standardised ($M = 0$, $SD = 1$) and modified (see below).

*Comparison choice.* Comparison choice was measured by asking students to nominate the classmate in each of the seven academic subjects with whom they preferred to compare their grades. Participants were instructed to leave these items blank if they did not compare with anyone. On average, across the seven subject areas 19% of the sample did not nominate a comparison person.

The comparison student’s grade at Time 2 was used to ascertain comparison direction. Table 5.1 presents the average scores for each academic subject for the comparison targets, as reported by Blanton et al. (1999). Blanton et al. conducted paired $t$ tests to ascertain whether the comparison person’s grades and the individual’s grades differed. Significant differences were demonstrated in all subjects, with the exception of geography and history. Blanton et al. concluded that those who did nominate a comparison other tended to compare slightly upward, that is, they chose someone who was performing slightly better in the course than they themselves were.

*Self-evaluation.* Self-evaluation was measured by asking students to rate their performance compared to their classmates in the seven academic subjects detailed above, using a 5-point Likert scale, which ranged from 1 (*much worse*) to 5 (*much better*), with a mid-point of 3 (*the same*). Blanton et al. (1999) provided evidence (e.g., the average inter-item correlation = .25) that these course-specific self-
evaluations were course-related appraisals and “not simply seven different ways of assessing the same general appraisal” (p. 424). The means and standard deviations for the seven subject-specific self-evaluations, as reported by Blanton et al., are presented in Table 5.1. This measure was standardised ($M = 0$, $SD = 1$) for subsequent analyses.


<table>
<thead>
<tr>
<th>Subject</th>
<th>Individual T2 Grade $M$</th>
<th>SD</th>
<th>Comparison Choice Grade $M$</th>
<th>SD</th>
<th>Self-evaluation $M$</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>7.04</td>
<td>1.21</td>
<td>7.15</td>
<td>1.24</td>
<td>3.07</td>
<td>0.72</td>
</tr>
<tr>
<td>Dutch</td>
<td>6.53</td>
<td>1.22</td>
<td>6.68</td>
<td>1.23</td>
<td>2.92</td>
<td>0.73</td>
</tr>
<tr>
<td>English</td>
<td>6.66</td>
<td>1.47</td>
<td>6.87</td>
<td>1.53</td>
<td>2.95</td>
<td>0.96</td>
</tr>
<tr>
<td>French</td>
<td>6.80</td>
<td>1.61</td>
<td>7.05</td>
<td>1.65</td>
<td>3.02</td>
<td>0.96</td>
</tr>
<tr>
<td>Geography</td>
<td>6.82</td>
<td>1.13</td>
<td>6.89</td>
<td>1.19</td>
<td>2.96</td>
<td>0.73</td>
</tr>
<tr>
<td>History</td>
<td>6.71</td>
<td>1.32</td>
<td>6.81</td>
<td>1.40</td>
<td>3.01</td>
<td>0.81</td>
</tr>
<tr>
<td>Math</td>
<td>6.61</td>
<td>1.48</td>
<td>6.85</td>
<td>1.44</td>
<td>3.10</td>
<td>0.87</td>
</tr>
</tbody>
</table>

**Modified grades.** Standardised academic achievement tests that provide comparable scores on a common metric for all students from different classes and schools are routinely used in BFLPE studies. However, in the absence of such a test, the current study used grades at Time 1 and Time 2 to assess performance. Unfortunately, teachers tend to grade-on-a-curve, that is, they give grades depending on their frame of reference, usually the intelligence level of the students in their classes. This grading-on-a-curve results in there being little variation between classes in terms of the average grade assigned, even when there are substantial differences between classes in terms of the ability levels of students within classes (e.g., Marsh, 1987). Thus, we have the situation in New South Wales, Australia, where grades given to students in their last year of high school are statistically modified to account for this kind of idiosyncratic grading, so that grades can be compared. The Dutch system is no different. Teachers in the Blanton et al. (1999) study were able to set their own tests and assign grades according to their own criteria, thus undoubtedly leading to the type of grading-on-a-curve effect just described. Because of a potential idiosyncrasy in absolute grade levels assigned by
different teachers, Blanton et al. removed the between-class effects by creating a set of 32 dummy variables. However, for purposes of evaluating the BFLPE, it was critical that there was a class-average grade that reflected the differing ability levels of the classes. Fortunately, information regarding the structure of the classes in the Blanton et al. study was provided by one of the authors (H. Kuyper, personal communication, April, 2004). This information, described in the following paragraphs, allowed classes to be scaled to reflect different ability levels.

At the time of the Blanton et al. (1999) study the Dutch educational system had five different tracks. The lowest two were “vocational” and the highest three were “general education”. The schools in the Blanton study only offered the three highest tracks. In the first year these tracks were mixed and were denoted by the terms “low”, “medium”, and “high”. Accordingly, there were four different types of classes in the Blanton study: low, low/medium, low/medium/high, and medium/high (H. Kuyper, personal communication, April, 2004). The mixture and number of class types across schools is shown in Table 5.2.

<table>
<thead>
<tr>
<th>Type of Class</th>
<th>School 1</th>
<th>School 2</th>
<th>School 3</th>
<th>School 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Low/Medium</td>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Low/Medium/High</td>
<td></td>
<td>5</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Medium/High</td>
<td>7</td>
<td></td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

For purposes of the present study, these categorisations were used to rescale the average ability levels on a 0-to-3 scale according to the class type, such that low = 0, medium = 1.5, and high = 3. As such, the low class type was assigned the value of 0, the medium class type the value of 1.5, and the high class type the value of 3. Thus the four class types were assigned values of low = 0, low/medium = .75 (i.e., \([0 + 1.5]/2\)), low/medium/high = 1.5 (i.e., \([0 + 1.5 +3]/3\)), and medium/high = 2.25 (i.e., \([1.5 + 3]/2\)). This constant value was then added to the (within-class) standardised grades of each student, depending on the class type the student attended. Thus, Table
5.3 indicates that students in the low track had a constant value of 0 added to their score, those in the low/medium track had a constant value of + .75 added, and students in the highest track had a constant value of + 2.25 added. The grand mean across all students was then set to zero by standardising these scaled values.

<table>
<thead>
<tr>
<th>Type of Class</th>
<th>Low</th>
<th>Low/Medium</th>
<th>Low/Medium /High</th>
<th>Medium/ High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0</td>
<td>.75</td>
<td>1.5</td>
<td>2.25</td>
</tr>
</tbody>
</table>

*Note. Constant added to the (within-class) standardised grades of each student.*

A critical issue in this transformation was how much difference there was likely to be in achievement levels of different class types. Two approaches were taken to this issue. First, the PISA 2003 database (see Organisation for Economic Cooperation and Development, 2005a, 2005b) provided information on which to establish the transformations. Using data from the Netherlands, the intraclass correlation (ICC) was evaluated. The ICC indicated the proportion of the total variance in math achievement scores that could be explained by differences between schools. Analyses showed the ICC for the Netherlands to be .59 (see Appendix B). Second, based on a total achievement score from the PISA 2000 database (see Organisation for Economic Cooperation and Development, 2001a, 2001b), reflecting mathematics, science, and literacy, the ICC was found to be .46 (see Marsh & Hau, 2003). In each case, the results indicated that there were large systematic differences between schools in the Netherlands in terms of school-average achievement levels. Whereas the real interest was the intra-class correlation, the amount of variance accounted for by differences between classes rather than schools, there was no class identification in the PISA data. However, it can be assumed that if there are large differences between schools due to tracking, there must also be corresponding differences between classes. Consequently, transforming the data so that the intraclass correlation was no more than .46 provided a conservative estimate of differences between class-average ability that actually existed in the Netherlands. Using the 0-to-3 scale described above, the intra-class correlation for the transformed
data in the current study was .35 (the results were similar for each school subject because the same transformation was used for scores in each school subject; see Appendix B).

Two additional pieces of evidence further strengthened the decision to transform these grades. First, the ICC across schools and classes was examined for the unmodified grades. For each subject the ICC = 0 across schools and classes, implying that there were no differences in T2 unmodified grades across the schools or across classes (see Appendix B). However, as these schools did have classes differing in ability, this suggests that Dutch teachers were grading-on-a-curve, awarding grades as a function of their frame of reference. Second, preliminary analyses from large-scale cohort studies currently being undertaken in Dutch high schools (H. Kuyper, personal communication, September 1, 2005) also support the transformation, by demonstrating that there are systematic differences between class types that are consistent with those posited here. In these studies, students in their first year of high school completed standardised achievement tests, consisting of 60 four-choice items, in Dutch, arithmetic, and information processing. Preliminary analyses demonstrated that mean scores on this test varied as a function of class type (i.e., low class type mean = 33.1; low/medium class type mean = 36.2; low/medium/high class type mean = 39.4; medium/high class type = 44.5). Hence, on the basis of this evidence, scaling classes in the manner adopted seemed appropriate. Thus, although somewhat arbitrary, and rather conservative, these transformations allowed the grading-on-a-curve effect to be eliminated and provided an approximate scale to compare students in different track types in a way that would not otherwise have been possible. Importantly, this transformation had no effect on the relative achievement scores of students within a given class.

**Research Design**

Blanton et al.’s (1999) study was undertaken across one academic school year. Grades were accessed on three occasions from trimester grade reports. At the end of trimester 2, participants completed a questionnaire containing the comparison choice and self-evaluation measures.
**Procedure and Statistical Analyses**

**Missing data.** As reported by Blanton et al. (1999), an average of 19% of the sample chose not to nominate a comparison target. The statistical programs used to analyse these data (SPSS v.14 and MLwiN v.2.02) routinely use listwise deletion of non-completers in generating results. However, as this study was a further analysis of these previously published data, the decision was made not to conduct imputations to compensate for the missing data.

**Standardisation.** All variables were standardised (z-scored) to have $M = 0$, $SD = 1$ across the entire sample (see Aiken & West, 1991; Marsh & Rowe, 1996; Raudenbush & Bryk, 2002). The original grades were modified as previously described and these modified school grades were then standardised ($M = 0$, $SD = 1$). School-average ability was determined by taking the average T2 modified grade for students in each class (but not restandardising these scores so that individual student and school-average ability scores were in the same metric).

**Replication of original results.** Regression analyses, using SPSS (v.14), were used to replicate Blanton et al.’s (1999) original results for grades and self-evaluation. When T2 grade was the outcome variable, predictors were T1 grade, self-evaluation, and comparison choice. When self-evaluation was the outcome variable, predictors were T2 grade and comparison choice.

However, as there was a multilevel structure in these data, multilevel modelling was used in all subsequent analyses. The multilevel structure was comprised of students within classes, and there were numerous classes within a school. This produced three levels, with individual students being level one, classes at level two, and schools at level three. Each lower level can be thought of as being grouped, or nested, within higher levels (refer to Chapter 4 for a detailed account of multilevel modelling). Hence, multilevel modelling was used, by means of the statistical program MLwiN (v.2.02), to ascertain whether Blanton et al.’s results could be replicated using this more powerful statistical tool, for both the unmoderated and the moderated grades.

Blanton et al.’s (1999) original analyses, based on unmoderated grades, were recreated for grades and self-evaluation, using multilevel modelling. Subsequently,
similar multilevel modelling analyses were conducted using the moderated grades, to demonstrate that Blanton et al.’s results could be replicated using these modified grades. For both these sets of analyses (using unmoderated and moderated grades), once again when the outcome variable was T2 grade, T1 grade, self-evaluation, and comparison choice were entered as predictors. When self-evaluation was the outcome variable, predictors were T2 grade and comparison choice. As these were replications of previous results, all effects used in the multilevel modelling analyses, except the intercept, were fixed.

**BFLPE analyses.** Multilevel modelling analyses were conducted to establish whether the BFLPE was present in Blanton et al.’s (1999) data, using the moderated grades. Two sets of analyses were conducted: one to demonstrate the moderating effect of comparison choice on the BFLPE, the other to demonstrate the moderating effect of individual ability on the BFLPE. In the first instance, self-evaluation for each subject was the outcome variable, and predictors were T1 moderated grades, class-average ability, comparison choice moderated grades, and the interaction of class-average ability and comparison choice moderated grades. As regards the moderating effect of individual ability, self-evaluation for each subject was the outcome variable and predictors were T1 moderated grades, class-average ability, and the interaction of class-average ability and individual ability as evidenced by T1 moderated grades. The product terms used to test the interaction effects were constructed by calculating the product of standardised (z-score) variables, but the product terms themselves were not restandardised. None of the effects (with the exception of the intercept) were allowed to be random as it was not necessary to do so to answer the hypotheses or research questions posited to investigate the aims of the current study.

Residuals were inspected at different levels for all BFLPE analyses. At higher levels in the Blanton et al. (1999) data there was slight evidence of non-linearity and heteroscedasticity for some academic subjects. However, the assumptions were not severely violated. As failures in the assumptions of linearity and homoscedasticity weaken rather than invalidate analyses (Tabachnick & Fidell, 2007), and as these analyses were of previously published data, no action was necessary.
**Preliminary Descriptive Statistics**

*Modified grades.* Means and standard deviations for unstandardised modified grades for individuals and their comparison choices are presented in Table 5.4. Also presented therein are the results of paired *t* tests comparing individual modified grades and comparison choice modified grades. These analyses demonstrated that modifying the grades made no difference to the pattern of results. Consistent with Blanton et al. (1999), participants chose to compare with a better performing other in each subject, although this was statistically significant for only five subjects, with geography and history not being statistically significant.

Table 5.4. Individual Grades, Comparison Choice Grades, and Paired *t* Tests for Unstandardised Modified Grades in Blanton et al.’s (1999) Data

<table>
<thead>
<tr>
<th>Subject</th>
<th>Individual T2 Grade</th>
<th>Comparison Choice Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>M</em></td>
<td><em>SD</em></td>
</tr>
<tr>
<td>Biology</td>
<td>1.58</td>
<td>1.24</td>
</tr>
<tr>
<td>Dutch</td>
<td>1.58</td>
<td>1.24</td>
</tr>
<tr>
<td>English</td>
<td>1.58</td>
<td>1.24</td>
</tr>
<tr>
<td>French</td>
<td>1.59</td>
<td>1.24</td>
</tr>
<tr>
<td>Geography</td>
<td>1.58</td>
<td>1.24</td>
</tr>
<tr>
<td>History</td>
<td>1.58</td>
<td>1.25</td>
</tr>
<tr>
<td>Math</td>
<td>1.56</td>
<td>1.24</td>
</tr>
</tbody>
</table>

*Note.* Unstandardised moderated grades = raw grades standardised with constant added. SDD = Standard deviation of the difference. df and *t* values are from paired *t* tests which were used to determine if there were statistical differences between an individual’s modified grade and comparison choice’s modified grade. A tendency toward upward comparison is indicated by negative *t* values. *p < .05. **p < .01.

**Results**

**Overview of Analyses**

To test the hypotheses and research questions posed to satisfy the aims of Study 2a, five sets of analyses were conducted. Firstly, Blanton et al.’s original results for grades and self-evaluation were replicated using multiple regression (Hypotheses 1.1 and 1.2). When the outcome variable was T2 grades, predictor variables were T1 grade, self-evaluation, and comparison choice grade. When the outcome variable was self-evaluation, predictors were T2 grade and comparison choice grade.
Then, Blanton et al.’s original results for grades and self-evaluation were replicated using multilevel modelling with the unmoderated and moderated grades (Hypotheses 2.1 and 2.2). In these two sets of analyses, when the outcome variable was T2 grade, predictors were T1 grade, self-evaluation, and the comparison choice grade. When self-evaluation was the outcome variable, T2 grade, and the comparison choice grade were entered as predictors.

Fourthly, multilevel modelling analyses were conducted with moderated grades to test for the presence of the BFLPE (Hypotheses 3.1 and 3.2) and to ascertain whether comparison choice moderated the BFLPE (Hypothesis 4.1 and Research Question 4.1). The outcome variable was self-evaluation for each subject, and predictors were T1 moderated grades, class-average ability, comparison choice moderated grades, and the interaction of class-average ability and comparison choice moderated grades.

Lastly, multilevel modelling analyses were conducted with moderated grades to ascertain if individual ability moderated the BFLPE (Research Question 5.1). The outcome variable was self-evaluation for each subject and T1 moderated grades, class-average ability, and the interaction of class-average ability and individual ability, as evidenced by T1 moderated grades, were entered as predictors.

**Results for Hypothesis 1.1: Replication of Blanton et al.’s (1999) Results Utilising Multiple Regression – Grades**

Hypothesis 1.1 predicted that the results of the Blanton et al. (1999) study would be replicated for grades whereby, based on multiple regression, for all seven school subjects T2 grade would be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice.

As indicated in Table 5.5, for all seven school subjects T2 grade was significantly positively predicted by T1 grade (ranging from 0.49 for history to 0.65 for French), T2 self-evaluation (ranging from 0.25 for Dutch and English to 0.37 for geography), and T2 comparison choice (ranging from 0.07 for English to 0.17 for geography). Given that Blanton et al.’s (1999) results were replicated for grades, Hypothesis 1.1 is accepted.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Blanton et al.’s original results</th>
<th>Recreation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade</td>
<td>T2 SE</td>
</tr>
<tr>
<td>Biology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.57</td>
<td>0.36</td>
</tr>
<tr>
<td>t</td>
<td>17.52**</td>
<td>7.97**</td>
</tr>
<tr>
<td>Dutch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.69</td>
<td>0.24</td>
</tr>
<tr>
<td>t</td>
<td>24.08**</td>
<td>6.04**</td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.63</td>
<td>0.35</td>
</tr>
<tr>
<td>t</td>
<td>22.5**</td>
<td>8.82**</td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.70</td>
<td>0.45</td>
</tr>
<tr>
<td>t</td>
<td>26.44**</td>
<td>11.71**</td>
</tr>
<tr>
<td>Geography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.59</td>
<td>0.38</td>
</tr>
<tr>
<td>T</td>
<td>20.43**</td>
<td>9.72**</td>
</tr>
<tr>
<td>History</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.55</td>
<td>0.44</td>
</tr>
<tr>
<td>t</td>
<td>15.5**</td>
<td>9.83**</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.67</td>
<td>0.32</td>
</tr>
<tr>
<td>T</td>
<td>21.69**</td>
<td>8.25**</td>
</tr>
</tbody>
</table>

**p < .01 Regression coefficients and degrees of freedom are slightly different in the recreation. Thirty-two dummy variables were used in the Blanton study to “centre effects within each classroom” (Blanton et al., 1999, p. 425), but were not used in the recreation.
**Results for Hypothesis 1.2: Replication of Blanton et al.’s (1999) Results Utilising Multiple Regression – Self-Evaluation**

This hypothesis predicted that the results of the Blanton et al. (1999) study would be replicated for self-evaluation. It was anticipated that, based on multiple regression, self-evaluation for all seven school subjects would be significantly positively predicted by T2 grade, but T2 comparison choice would not be a significant predictor of self-evaluation for any of the academic subjects.

Results regarding the replication of Blanton et al.’s (1999) results for self-evaluation using multiple regression analyses are presented in Table 5.6. In all seven school subjects, self-evaluation was significantly positively predicted by T2 grade, ranging from 0.31 for biology to 0.65 for French. The relation between comparison choice and self-evaluation was not significant for any of the school subjects tested. As Blanton et al.’s (1999) results were replicated for self-evaluation, Hypothesis 1.2 is accepted.

**Summary for Hypotheses 1.1 and 1.2: Replication of Blanton et al.’s (1999) Results Utilising Multiple Regression – Grades and Self-Evaluation**

Using multiple regression Blanton et al.’s (1999) findings were replicated for grades and self-evaluation. When T2 grade was the outcome variable, results demonstrated that for each course previous ability, a positive evaluation of a student’s own ability in that course, and the comparison with a better performing target were predictive of higher grades at T2. As regards self-evaluation, ability in a particular subject was the most important determinant of a positive evaluation of a student’s own ability in that course. Comparisons with a selected individual were not associated with self-evaluations.
Table 5.6. Recreation of Blanton et al.’s (1999) Data, Regressing T2 Self-Evaluation on T2 Grade and T2 Comparison Choice, Using Non-Standardised Regression Coefficients

<table>
<thead>
<tr>
<th>Subject</th>
<th>Blanton et al.’s original results</th>
<th>Recreation,</th>
<th>df</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T2 Grade</td>
<td>T2 Choice</td>
<td>T2 Grade</td>
<td>T2 Choice</td>
</tr>
<tr>
<td>Biology</td>
<td>0.29</td>
<td>0.04</td>
<td>0.31</td>
<td>0.05</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>11.73**</td>
<td>1.68</td>
<td>11.69**</td>
<td>1.95</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.32</td>
<td>0.01</td>
<td>0.35</td>
<td>-0.00</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>12.24**</td>
<td>0.25</td>
<td>12.67**</td>
<td>-0.07</td>
</tr>
<tr>
<td>English</td>
<td>0.44</td>
<td>-0.02</td>
<td>0.62</td>
<td>-0.01</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>20.15**</td>
<td>0.79</td>
<td>19.85**</td>
<td>-0.22</td>
</tr>
<tr>
<td>French</td>
<td>0.44</td>
<td>-0.02</td>
<td>0.65</td>
<td>-0.03</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>22.53**</td>
<td>1.03</td>
<td>22.50**</td>
<td>-0.97</td>
</tr>
<tr>
<td>Geography</td>
<td>0.35</td>
<td>0.01</td>
<td>0.37</td>
<td>0.02</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>13.35**</td>
<td>0.56</td>
<td>13.90**</td>
<td>0.71</td>
</tr>
<tr>
<td>History</td>
<td>0.38</td>
<td>-0.03</td>
<td>0.46</td>
<td>-0.01</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>15.81**</td>
<td>1.13</td>
<td>15.94**</td>
<td>-0.56</td>
</tr>
<tr>
<td>Math</td>
<td>0.41</td>
<td>-0.02</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>16.63**</td>
<td>-0.94</td>
<td>16.05**</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

**p < .01 , Regression coefficients and degrees of freedom are slightly different in the recreation. Thirty-two dummy variables were used in the Blanton study to “centre effects within each classroom” (Blanton et al., 1999, p. 425), but were not used in the recreation.
Results for Hypothesis 2.1: Replication of Blanton et al.’s (1999) Results Utilising Multilevel Modelling – Grades

Hypothesis 2.1 predicted that Blanton et al.’s (1999) results would be replicated for grades in all seven school subjects when multilevel modelling was used. It was anticipated that T2 grade would be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice.

Table 5.7 presents the results for the replication of Blanton et al.’s (1999) findings for unmoderated grades for all seven subjects using multilevel modelling. Using unmoderated grades, T2 grade was significantly positively predicted by T1 grade (ranging from 0.487 for history to 0.645 for French), T2 self-evaluation (ranging from 0.182 for Dutch to 0.291 for French), and T2 comparison choice (ranging from 0.069 in English to 0.172 for geography).

Table 5.7. Recreation of Blanton et al.’s (1999) Original Results for Grades, Using Multilevel Modelling and Unmoderated Grades

<table>
<thead>
<tr>
<th>Subject</th>
<th>T1 Grade (SE)</th>
<th>T2 Self-Evaluation (SE)</th>
<th>T2 Choice (SE)</th>
<th>Cons (SE)</th>
<th>School (SE)</th>
<th>Class (SE)</th>
<th>Individual (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>0.502*</td>
<td>0.252*</td>
<td>0.079*</td>
<td>0.015</td>
<td>0.000</td>
<td>0.000</td>
<td>0.544*</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.635*</td>
<td>0.182*</td>
<td>0.088*</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
<td>0.385*</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.025)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>English</td>
<td>0.621*</td>
<td>0.243*</td>
<td>0.069*</td>
<td>-0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>0.303*</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>French</td>
<td>0.645*</td>
<td>0.291*</td>
<td>0.088*</td>
<td>-0.014</td>
<td>0.000</td>
<td>0.000</td>
<td>0.231*</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Geography</td>
<td>0.534*</td>
<td>0.272*</td>
<td>0.172*</td>
<td>0.006</td>
<td>0.000</td>
<td>0.000</td>
<td>0.384*</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.027)</td>
<td>(0.025)</td>
<td>(0.025)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>History</td>
<td>0.487*</td>
<td>0.286*</td>
<td>0.141*</td>
<td>-0.016</td>
<td>0.000</td>
<td>0.000</td>
<td>0.437*</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.029)</td>
<td>(0.025)</td>
<td>(0.026)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Math</td>
<td>0.576*</td>
<td>0.236*</td>
<td>0.116*</td>
<td>-0.005</td>
<td>0.000</td>
<td>0.000</td>
<td>0.389*</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.028)</td>
<td>(0.025)</td>
<td>(0.025)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.021)</td>
</tr>
</tbody>
</table>

Note. All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. T2 Unmoderated Grade is the outcome variable. T1 Grade (unmoderated), T2 Self-evaluation, and T2 Comparison Choice (unmoderated) are predictor variables. Cons = Constant. Standard Error (SE) in parentheses.
These multilevel modelling analyses were also conducted with the modified grades. T2 moderated grade was significantly positively predicted by T1 moderated grade (ranging from 0.560 for history to 0.710 for French), T2 self-evaluation (ranging from 0.122 for Dutch to 0.200 for French), and T2 moderated comparison choice (ranging from 0.123 for English to 0.232 for geography). Table 5.8 presents details for these moderated grades.

Using the more powerful statistical method of multilevel modelling, Blanton et al.’s (1999) results for grades were replicated for both unmoderated and moderated grades. Whether the grades were moderated or not, for each subject, previous ability in that subject, a positive self-evaluation, and comparison with a better performing other were associated with improved T2 grades. Hence, Hypothesis 2.1 is accepted.

Table 5.8. Recreation of Blanton et al.’s (1999) Original Results for Grades, Using Multilevel Modelling and Moderated Grades

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade (moderated) (SE)</td>
<td>T2 Self-Eval (moderated) (SE)</td>
</tr>
<tr>
<td>Biology</td>
<td>0.577* (0.031)</td>
<td>0.180* (0.025)</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.700* (0.025)</td>
<td>0.122* (0.021)</td>
</tr>
<tr>
<td>English</td>
<td>0.705* (0.025)</td>
<td>0.154* (0.020)</td>
</tr>
<tr>
<td>French</td>
<td>0.710* (0.022)</td>
<td>0.200* (0.018)</td>
</tr>
<tr>
<td>Geography</td>
<td>0.617* (0.025)</td>
<td>0.180* (0.021)</td>
</tr>
<tr>
<td>History</td>
<td>0.560* (0.030)</td>
<td>0.198* (0.023)</td>
</tr>
<tr>
<td>Math</td>
<td>0.672* (0.026)</td>
<td>0.151* (0.022)</td>
</tr>
</tbody>
</table>

Note. All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. T2 Self-Eval = T2 Self-evaluation; Cons = Constant. Moderated T2 Grade is the outcome variable. Moderated T1 Grade, T2 Self-evaluation, and T2 Comparison Choice (moderated) are predictor variables. Standard Error (SE) in parentheses.
Results for Hypothesis 2.2: Replication of Blanton et al.’s (1999) Results Utilising Multilevel Modelling – Self-Evaluation

This hypothesis predicted that Blanton et al.’s (1999) results would be replicated for self-evaluation in all seven school subjects when multilevel modelling was used in place of multiple regression. It was anticipated that self-evaluation would be significantly positively predicted by T2 grade but T2 comparison choice would not be a significant predictor of self-evaluation.

Using multilevel modelling and unmoderated grades, T2 grade positively predicted self-evaluation in each subject, ranging from 0.429 for biology to 0.683 for French (see Table 5.9). The effect of T2 comparison choice on self-evaluation was not statistically significant for six of the seven subjects. Comparison choice was a statistically significant positive predictor of self-evaluation for biology.

Using multilevel modelling and moderated grades, a somewhat similar pattern of results emerged (see Table 5.10). T2 grade was a statistically significant positive predictor of self-evaluation in each subject, ranging from 0.478 for biology to 0.815 for French. The effect of T2 comparison choice on self-evaluation was not statistically significant for any of the seven subjects.

Using multilevel modelling Blanton et al.’s (1999) results for self-evaluation were partially replicated for the unmoderated grades and fully replicated for the moderated grades. Whether the grades were moderated or not, for each subject, self-evaluation was predicted by ability in that subject. The selected comparison was not related to self-evaluation, with the exception of biology using the unmoderated grades, a result not found in the multiple regression analyses (see Hypothesis 1.2 and Table 5.6) or in the multilevel modelling analyses using moderated grades. Hence, Hypothesis 2.2 is not accepted for unmoderated grades in biology using multilevel modelling, as results for that subject were inconsistent with those of the original findings. However, Hypothesis 2.2 is accepted for all other subjects using the unmoderated grades. Hypothesis 2.2 is also accepted for moderated grades using multilevel modelling, as they were fully consistent with the original findings.
Table 5.9. *Recreation of Blanton et al.’s (1999) Original Results for Self-Evaluation using Multilevel Modelling and Unmoderated Grades*

<table>
<thead>
<tr>
<th>Subject</th>
<th>T2 Grade (SE)</th>
<th>T2 Choice (SE)</th>
<th>Constant (SE)</th>
<th>School (SE)</th>
<th>Class (SE)</th>
<th>Individual (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>0.429* (0.036)</td>
<td>0.071* (0.035)</td>
<td>-0.017 (0.093)</td>
<td>0.028 (0.024)</td>
<td>0.010 (0.013)</td>
<td>0.779* (0.043)</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.438* (0.038)</td>
<td>0.005 (0.037)</td>
<td>-0.003 (0.050)</td>
<td>0.003 (0.007)</td>
<td>0.016 (0.015)</td>
<td>0.798* (0.045)</td>
</tr>
<tr>
<td>English</td>
<td>0.653* (0.032)</td>
<td>-0.010 (0.030)</td>
<td>0.029 (0.065)</td>
<td>0.010 (0.012)</td>
<td>0.031* (0.015)</td>
<td>0.575* (0.032)</td>
</tr>
<tr>
<td>French</td>
<td>0.683* (0.030)</td>
<td>-0.028 (0.028)</td>
<td>0.004 (0.032)</td>
<td>0.001 (0.003)</td>
<td>0.002 (0.007)</td>
<td>0.515* (0.029)</td>
</tr>
<tr>
<td>Geography</td>
<td>0.511* (0.037)</td>
<td>0.025 (0.034)</td>
<td>-0.016 (0.037)</td>
<td>0.000 (0.000)</td>
<td>0.008 (0.011)</td>
<td>0.745* (0.041)</td>
</tr>
<tr>
<td>History</td>
<td>0.576* (0.036)</td>
<td>-0.020 (0.033)</td>
<td>-0.005 (0.041)</td>
<td>0.001 (0.005)</td>
<td>0.012 (0.012)</td>
<td>0.704* (0.039)</td>
</tr>
<tr>
<td>Math</td>
<td>0.563* (0.034)</td>
<td>0.008 (0.033)</td>
<td>-0.024 (0.095)</td>
<td>0.029 (0.025)</td>
<td>0.028 (0.016)</td>
<td>0.656* (0.037)</td>
</tr>
</tbody>
</table>

*Note.* All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. T2 Self-evaluation is the outcome variable. T2 Grade (unmoderated) and T2 Comparison Choice (unmoderated) are predictors. Standard Error (SE) in parentheses.

**Summary for Hypotheses 2.1 and 2.2: Replication of Blanton et al.’s (1999)**

**Results Utilising Multilevel Modelling – Grades and Self-Evaluation**

Using multilevel modelling, Blanton et al.’s (1999) findings for grades were replicated for both unmoderated and moderated grades. When self-evaluation was the outcome variable, using multilevel modelling Blanton et al.’s (1999) findings were replicated for all academic subjects when moderated grades were analysed, and for all academic subjects with the exception of biology when the unmoderated grades were analysed.
Table 5.10. *Recreation of Blanton et al.’s (1999) Original Results for Self-Evaluation using Multilevel Modelling and Moderated Grades*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T2 Grade (moderated)</td>
<td>Constant (SE)</td>
</tr>
<tr>
<td>Biology</td>
<td>0.478* (0.044)</td>
<td>-0.015 (0.079)</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.540* (0.045)</td>
<td>0.000 (0.065)</td>
</tr>
<tr>
<td>English</td>
<td>0.772* (0.039)</td>
<td>0.021 (0.093)</td>
</tr>
<tr>
<td>French</td>
<td>0.815* (0.037)</td>
<td>-0.005 (0.085)</td>
</tr>
<tr>
<td>Geography</td>
<td>0.578* (0.045)</td>
<td>-0.015 (0.074)</td>
</tr>
<tr>
<td>History</td>
<td>0.665* (0.044)</td>
<td>-0.002 (0.079)</td>
</tr>
<tr>
<td>Math</td>
<td>0.653* (0.042)</td>
<td>-0.030 (0.084)</td>
</tr>
</tbody>
</table>

*Note.* All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. T2 Self-evaluation is the outcome variable. Moderated T2 Grades and T2 Comparison Choice (moderated) are predictors. Standard Error (SE) in parentheses.


Hypothesis 3.1 predicted that, using multilevel modelling, for each school subject in Blanton et al.’s (1999) data, self-evaluation would be significantly positively predicted by individual ability as measured by T1 grade. Results demonstrated that for all seven school subjects in Blanton et al.’s (1999) data, individual ability was a significant positive predictor of self-evaluation. As presented in Table 5.11, effects varied from .389 for biology to .735 for English. Given that there was a significantly positive relation between individual ability and self-evaluation, Hypothesis 3.1 is supported.

Hypothesis 3.2 predicted that, using multilevel modelling, for each school subject in Blanton et al.’s (1999) data, self-evaluation would be significantly negatively predicted by class-average ability, as measured by T2 class-average grade. Results indicated that class-average ability was a statistically significant negative predictor of self-evaluation for all seven school subjects. These effects varied from -.356 for Dutch to -.603 for English (see Table 5.11). Consequently, as self-evaluation was negatively predicted by class-average ability in all school subjects, Hypothesis 3.2 is accepted.


Results for Hypotheses 3.1 and 3.2 indicate that a BFLPE was present in these data. While individual ability was statistically significantly positively related to self-evaluation, class-average ability was always significantly negatively related. In view of the fact that it is this negative association that characterises the BFLPE, it is evident that the BFLPE is present in Blanton et al.’s (1999) data.


This hypothesis anticipated that, for each school subject, the comparison person’s grade would not have a statistically significant association with self-evaluation in Blanton et al.’s (1999) data. Results of the analyses showed that, for all seven subjects, the relation between the comparison person’s ability and self-evaluation was not statistically significant (see Table 5.11). These results are consistent with the original study and demonstrate that comparing with others was not related to self-evaluations of ability. Given that comparison choice was not significantly related to self-evaluation, Hypothesis 4.1 is accepted.

Research Question 4.1 enquired whether, for each school subject, the comparison person’s ability could moderate the effect of class-average ability on self-evaluation in Blanton et al.’s (1999) data, or whether the BFLPE co-existed with the positive effects of upward comparison on performance.

Multilevel modelling analyses of Blanton et al.’s (1999) data demonstrated that, for six of the seven individual school subjects, the class-average ability X comparison choice interaction did not have a statistically significant association with self-evaluation (see Table 5.11). However, this interaction had a significant positive relation with self-evaluation for Dutch (.116). Figure 5.2 demonstrates the pattern of this interaction. Smaller BFLPEs were associated with students who compared with high-performing others. Students in low-ability classes who compared with low-performing others had slightly higher self-evaluations for Dutch than those who compared themselves with others whose performances were average or above average. In high-ability classes this pattern was reversed. In high-ability classes, those who compared themselves with high-performing others had higher self-evaluations for Dutch than those who compared themselves with students whose performances were average or below average.

Summary for Hypothesis 4.1 and Research Question 4.1: Comparison Choice

The main effect of comparison choice on self-evaluation was not significant for any of the academic subjects tested. Hence, choosing to compare with another student had no effect on a student’s self-evaluation. Additionally, the current study demonstrated that the comparison person’s grade did not moderate the effect of class-average ability on self-evaluation (i.e., the BFLPE) for six of the seven academic subjects tested. Thus, for those six subjects, no matter with whom a student chose to compare him/herself with, the higher the ability of the class, the lower was the student’s self-evaluation. Results, however, indicated that the BFLPE was moderated for Dutch, whereby the BFLPE was reduced for students who chose to compare themselves with a high-performing other.
Table 5.11. *Multilevel Modelling Analyses of Blanton et al.’s (1999) Data Predicting Self-Evaluation, with Comparison Choice as a Moderator of the BFLPE*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade</td>
<td>Class-average Ability</td>
</tr>
<tr>
<td>Biology</td>
<td>0.389*</td>
<td>-0.429*</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.492*</td>
<td>-0.356*</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.080)</td>
</tr>
<tr>
<td>English</td>
<td>0.735*</td>
<td>-0.603*</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.074)</td>
</tr>
<tr>
<td>French</td>
<td>0.701*</td>
<td>-0.577*</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Geography</td>
<td>0.440*</td>
<td>-0.466*</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>History</td>
<td>0.590*</td>
<td>-0.538*</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>Math</td>
<td>0.625*</td>
<td>-0.477*</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.072)</td>
</tr>
</tbody>
</table>

*Note.* Standard Error in Parentheses (SE). All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores.
Research question 5.1 enquired about the nature of the relation between the individual ability by class-average ability interaction and self-evaluation for each school subject in Blanton et al.’s (1999) data. When individual ability was considered as a moderator of the BFLPE in Blanton et al.’s (1999) data, a BFLPE emerged. For all school subjects, individual ability was a significant positive predictor of self-evaluation (ranging from .387 for biology to .761 for English) and class-average ability was a significant negative predictor (ranging from -.355 for biology to -.632 for English; see Table 5.12). However, ability did not moderate the BFLPE, as the ability-by-class-average interaction was not significant for any of the seven school subjects tested. Students of all ability levels in Blanton et al.’s (1999) study suffered the negative effects of the BFLPE.

Figure 5.2. Comparison Other by Class-Average Ability Interaction for Dutch in Blanton et al. (1999)

Note. Based on predicted values. High class-average ability = 1 standard deviation above the mean for class-average Dutch ability and low class-average ability = 1 standard deviation below the mean. Similarly, high comparison other grade = 1 standard deviation above the mean for comparison other grade, and low comparison other grade = 1 standard deviation below the mean. Individual ability is held constant.

Results for Research Question 5.1: Individual Ability as a Moderator of the BFLPE in Blanton et al.’s (1999) Data Based on Multilevel Modelling

Research question 5.1 enquired about the nature of the relation between the individual ability by class-average ability interaction and self-evaluation for each school subject in Blanton et al.’s (1999) data. When individual ability was considered as a moderator of the BFLPE in Blanton et al.’s (1999) data, a BFLPE emerged. For all school subjects, individual ability was a significant positive predictor of self-evaluation (ranging from .387 for biology to .761 for English) and class-average ability was a significant negative predictor (ranging from -.355 for biology to -.632 for English; see Table 5.12). However, ability did not moderate the BFLPE, as the ability-by-class-average interaction was not significant for any of the seven school subjects tested. Students of all ability levels in Blanton et al.’s (1999) study suffered the negative effects of the BFLPE.
Ability as a Moderator of the BFLPE*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade</td>
<td>Class-average Ability</td>
</tr>
<tr>
<td>Biology</td>
<td>0.387* (0.093)</td>
<td>-0.355* (0.069)</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.407* (0.084)</td>
<td>-0.364* (0.065)</td>
</tr>
<tr>
<td>English</td>
<td>0.761* (0.077)</td>
<td>-0.632* (0.062)</td>
</tr>
<tr>
<td>French</td>
<td>0.691* (0.074)</td>
<td>-0.577* (0.053)</td>
</tr>
<tr>
<td>Geography</td>
<td>0.461* (0.088)</td>
<td>-0.426* (0.065)</td>
</tr>
<tr>
<td>History</td>
<td>0.634* (0.081)</td>
<td>-0.478* (0.058)</td>
</tr>
<tr>
<td>Math</td>
<td>0.552* (0.082)</td>
<td>-0.418* (0.061)</td>
</tr>
</tbody>
</table>

*Note.* Standard Error in Parentheses (SE). All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores.
Study 2a Summary

This section has described Study 2a, a further analysis of Blanton et al.’s (1999) data with the principal aim of ascertaining whether individual selected comparisons moderated or co-existed with the BFLPE. This section has provided specific aims, hypotheses and research questions, rationales for the hypotheses and research questions, methodology, and results for Study 2a. Using multiple regression, Blanton et al.’s (1999) original results were replicated. With one exception, they were also replicated using multilevel modelling for both the unmoderated and moderated grades. Although the BFLPE was evident in these data, it was not moderated by individual ability in any school subject or by comparison choice for six of the seven academic subjects tested. The BFLPE was moderated for Dutch, whereby smaller BFLPEs were associated with students who compared with a high-performing other. The following section describes Study 2b, a further analysis of Huguet et al.’s (2001) data.

Study 2b: Further Analysis of Huguet et al. (2001)

Aims

The main aim of Study 2b was to further analyse the original Huguet et al. (2001) data and an expanded data set, collected at the same time as the original data, to determine if individual selected comparisons, associated with enhanced performance, moderate the BFLPE, or co-exist with the BFLPE. More specifically, Study 2b aimed to:

1. Analyse Huguet et al.’s (2001) results for performance and self-evaluation based upon multiple regression in order to replicate their original results;
2. Utilise multiple regression to extend the external validity of Huguet et al.’s results for performance and self-evaluation to an expanded sample;
3. Utilise multilevel modelling to analyse Huguet et al.’s original results for performance and self-evaluation in order to demonstrate that their original results could be replicated using this more appropriate statistical technique;
4. Analyse Huguet et al.’s expanded data to ascertain if the original results for performance and self-evaluation can be replicated using multilevel modelling;

5. Conduct further analyses of Huguet et al.’s expanded data utilising multilevel modelling to test for the presence of the BFLPE;

6. Utilise Huguet et al.’s expanded data to ascertain whether the BFLPE is moderated by upward comparisons or whether the BFLPE co-exists with the positive effects of comparisons on performance; and

7. Explore whether the BFLPE is moderated by individual student ability in Huguet et al.’s expanded data in order to explicate whether the BFLPE varies as a function of individual ability.

**Statement of Hypotheses and Research Questions**

**Hypothesis 1.1: Replication of Huguet et al.’s (2001) Results Utilising Multiple Regression – Grades**

The results of the Huguet et al. (2001) study will be replicated whereby, based on multiple regression, for all seven school subjects (biology, Dutch, English, French, geography, history, math) T2 grade will be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice 1. Additionally, T2 grade in French, math, drawing, and technology, will be significantly positively predicted by T2 comparison choice 2.

**Hypothesis 1.2: Replication of Huguet et al.’s (2001) Results Utilising Multiple Regression – Self-Evaluation**

The results of the Huguet et al. (2001) study will be replicated whereby, based on multiple regression, self-evaluation for all seven subjects will be significantly positively predicted by T2 grade but T2 comparison choice 1 will not be a significant predictor of self-evaluation. T2 comparison choice 2 will only be a significantly positive predictor of self-evaluation for biology; for all other subjects the effect of T2 comparison choice 2 will not be statistically significant.
Hypothesis 2.1: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multiple Regression – Grades

The results of the Huguet et al. (2001) study will be replicated for the expanded data, based on three school subjects (French, math, history/geography). It is predicted that, using multiple regression, for each school subject, T2 grade will be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice 1. Additionally, T2 grade in French and math (but not history/geography) will be significantly positively predicted by T2 comparison choice 2.

Hypothesis 2.2: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multiple Regression – Self-Evaluation

The results of the Huguet et al. (2001) study will be replicated for the expanded data, based on three school subjects (French, math, history/geography). It is predicted that, using multiple regression, for each school subject, self-evaluation will be significantly positively predicted by T2 grade, but T2 comparison choices 1 and 2 will not have a significant association with self-evaluation.


Using multilevel modelling, Huguet et al.’s (2001) results will be replicated for all seven school subjects, whereby T2 grade will be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice 1. Additionally, T2 comparison choice 2 will significantly positively predict T2 grades in French, math, drawing, and technology, but will not have a significant association with T2 grades in biology, English, and history/geography.


Using multilevel modelling, Huguet et al.’s (2001) results will be replicated for all seven school subjects, whereby self-evaluation will be significantly positively predicted by T2 grade, but the effect of comparison choice 1 will not be statistically significant. Whereas for French, English, history/geography, math, drawing and technology, T2 comparison choice 2 will not be a significant predictor of self-
evaluation, T2 comparison choice 2 will be a significant positive predictor of self-evaluation for biology.

**Hypothesis 4.1: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multilevel Modelling – Grades**

Using multilevel modelling, Huguet et al.’s (2001) results will be replicated for the expanded data based on three school subjects (French, math, and history/geography), whereby for each school subject T2 grade will be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice 1. Additionally, T2 grade in French and math (but not history/geography) will be significantly positively predicted by T2 comparison choice 2.

**Hypothesis 4.2: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multilevel Modelling – Self-Evaluation**

Using multilevel modelling, Huguet et al.’s (2001) results will be replicated for the expanded data based on three school subjects (French, math, history/geography), whereby for each school subject self-evaluation will be significantly positively predicted by T2 grade, but T2 comparison choices 1 and 2 will not be significant predictors of self-evaluation.


When Huguet et al.’s expanded data are analysed using multilevel modelling, it is predicted that for each school subject (French, math, history/geography), self-evaluation, as measured in the Huguet et al. (2001) study, will be significantly positively predicted by individual ability as measured by T1 grade. This relation is depicted in Figure 5.1.

It is predicted that when Huguet et al.’s expanded data are analysed using multilevel modelling, for each school subject (French, math, history/geography), self-evaluation as measured in the Huguet et al. (2001) study will be significantly negatively predicted by class-average ability, as measured by T2 class-average grade – the BFLPE (see Figure 5.1).


In the Huguet et al. expanded data, for each school subject, it is expected that the main effect of the comparison person’s grade, as measured by his or her T2 grade, on self-evaluation, as measured in the Huguet et al. (2001) study (see Figure 5.1), will not be statistically significant.


Does comparison choice moderate the BFLPE? More specifically, for each school subject in Huguet et al.’s expanded data, does the comparison person’s grade, as measured by his or her T2 grade, moderate the effect of class-average ability, as measured by T2 class-average grade, on self-evaluation as measured by Huguet et al. (2001)? If there is no moderation, then do the positive effects of comparison choice on performance and the negative effects of class-average ability on self-evaluation co-exist?

Research Question 7.1: Individual Ability as a Moderator of the BFLPE in Huguet et al.’s (2001) Expanded Data Based on Multilevel Modelling

Is the BFLPE moderated by individual ability level? For each school subject in Huguet et al.’s expanded data, what is the relation between the individual ability (as measured by T1 grade) X class-average ability (as measured by T2 class-average grade) interaction and self-evaluation (as measured in the Huguet et al. (2001) study)?
**Rationale for Hypotheses and Research Questions**

**Rationale for Hypotheses 1.1 and 1.2: Replication of Huguet et al. (2001) Utilising Multiple Regression – Grades and Self-Evaluation**

As these components of the current investigation used the same statistical technique as the original study, the same pattern of results is anticipated.

**Rationale for Hypothesis 2.1 and 2.2: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multiple Regression – Grades and Self-Evaluation**

Although the expanded data contain more cases, it is anticipated that the results will be similar to those of the original analyses.


As noted in the rationale for Hypotheses 2.1 and 2.2 in Study 2a, when data have a multilevel structure, multilevel modelling is a more appropriate statistical tool than traditional single level techniques (see Chapter 4). The original study on which this re-analysis was based used single level multiple regressions. However, as the results from these original analyses were strong, it is anticipated that similar results will be demonstrated in the current study when multilevel regression analyses are used instead of single level regression analyses.


The rationale for Hypothesis 3.1 for Study 2a noted that BFLPE theory and empirical research propose that individual ability is positively related to academic self-concept, (e.g., Craven et al., 2000; Marsh, 2005; Marsh & Craven, 2002; Marsh & Hau, 2003; Marsh et al., 2001; Marsh et al., in press; Mulkey et al., 2005). As was the case for Study 2a, academic self-concept per se was not measured in Study 2b. However, the self-evaluation measure used was considered to be an approximate measure of academic self-concept (see Methodology section below). Accordingly, it
is expected that the same relation that exists between individual ability and academic self-concept will also be found between individual ability and the single measure of self-evaluation used in the current study.


As noted in the rationale for Hypothesis 3.2 for Study 2a, the relation between class-average ability and academic self-concept has consistently been shown to be significantly negative (e.g., Craven et al., 2000; Marsh, 2005; Marsh & Craven, 2002; Marsh & Hau, 2003; Marsh et al., 2001; Marsh et al., in press; Mulkey et al., 2005). Although in the absence of an academic self-concept measure a single measure of self-evaluation was used in the current study, it is expected that a similar negative relation will be found in the current study.


With one exception, the original Huguet et al. (2001) study found that there was no relation between the comparison person’s grade and self-evaluation. As this exception pertained to a second comparison choice for one school subject only, and as the current study is evaluating the first comparison choice, it is anticipated that the comparison person’s grade will not be statistically significantly related to self-evaluation.


The rationale for Research Question 4.1 in Study 2a noted that no previous research of which the author is aware has examined the relation between a comparison choice-by-class-average ability interaction and self-evaluation. It was also suggested that the positive effects of upward comparison on performance might either moderate, or co-exist with the BFLPE. A research question is therefore posed
to illuminate these questions as past research has been unable to provide guidance for formulating clear predictions.

**Rationale for Research Question 7.1: Individual Ability as a Moderator of the BFLPE in Huguet et al.’s (2001) Expanded Data Based on Multilevel Modelling**

The rationale for Research Question 5.1 in Study 2a noted that there is conflicting evidence regarding the moderating effect of individual ability on the BFLPE. Consequently, a research question is posed to illuminate this issue.

**Methodology**

**Participants**

In France, students begin high school when they are 11 years old. For the last three years of high school, students are segregated on the basis of academic ability (see Chapter 2). As was the case in the Blanton et al. (1999) study, there was no random assignment of students to classes, and there may have been many reasons for a student choosing to attend a particular school.

*Original data.* The total number of students who participated in the original Huguet et al. (2001) study were 264 (135 male, 129 female), from 11 classes across two French high schools, ranging in age from 12 to 14 years. Each school contained 5 to 6 classes, and there were 22 to 26 students in each class. Students stayed in same class with the same students for the entire year.

*Expanded data.* The larger sample, which was part of a much larger unpublished data set, consisted of 1,156 students (537 male, 619 female) from 51 classes across 12 French high schools. Participants were in first to fifth grade of high school, with a mean age of 13.5 years. Each class contained 13 to 33 students, and there were 1 to 9 classes in a school.
Measures

Grades. For both the original data and the expanded data, grades were given on a 20-point scale with 20 representing a high grade. Descriptors relating to these grades were “extremely poor from 0 to 5, poor from 6 to 9, passable to satisfying from 10 to 14, and very satisfying from 15 to 20” (Huguet et al., 2001, p. 561). In the original data the courses comprised biology, drawing, English, French, history, math, and technology. Average T2 scores for the subjects of the original data are presented in Table 5.13. In the expanded data, participants took three school subjects in common: French, history/geography combined, and math. Grades in the original and expanded data sets were subsequently standardised ($M = 0, SD = 1$).

Comparison choice. Comparison choice was measured as in the Blanton et al. (1999) study. However, in the Huguet et al. (2001) study, students could nominate two comparison targets. Participants were instructed to leave these items blank if they did not compare themselves with anyone. On average across all courses 7.37% of the original sample and 12.46% of the extended sample did not nominate a target for comparison choice 1.

The comparison student’s grade at Time 2 was used to ascertain comparison direction. Table 5.13 presents the average scores for each academic subject for comparison choice 1 in the original data. Huguet et al. (2001) noted that those who did nominate a comparison other tended to compare slightly upward, that is, they chose someone who was performing slightly better in the course than themselves. On the basis of paired $t$ tests these upward comparisons were shown to be statistically significant for comparison choice 1 in all seven academic subjects of the original data.

Self-evaluation. Self-evaluation was measured as in the Blanton et al. (1999) study. Huguet et al. (2001) reported that, consistent with the Blanton et al. (1999) study, self-evaluation scores tended to cluster around the midpoint of 3 (no better or worse than others in class). Students in the original Huguet et al. study were more positively biased with regards to drawing ($M = 3.27, SD = 0.88$) and technology ($M = 3.29, SD = 0.73$). Means and standard deviations for this measure in the original data are presented in Table 5.13. These self-evaluation scores were subsequently standardised ($M = 0, SD = 1$).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Individual T2 Grade</th>
<th>Comparison Choice 1 Grade</th>
<th>Self-evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Biology</td>
<td>13.14</td>
<td>3.49</td>
<td>13.92</td>
</tr>
<tr>
<td>French</td>
<td>12.07</td>
<td>3.12</td>
<td>13.02</td>
</tr>
<tr>
<td>English</td>
<td>13.14</td>
<td>3.61</td>
<td>14.09</td>
</tr>
<tr>
<td>History/Geography</td>
<td>13.36</td>
<td>3.65</td>
<td>14.35</td>
</tr>
<tr>
<td>Math</td>
<td>12.84</td>
<td>3.34</td>
<td>13.82</td>
</tr>
<tr>
<td>Drawing</td>
<td>14.28</td>
<td>3.26</td>
<td>15.10</td>
</tr>
<tr>
<td>Technology</td>
<td>13.86</td>
<td>3.09</td>
<td>14.51</td>
</tr>
</tbody>
</table>

*Class-average ability.* School grades in the French system are specifically designed to be comparable across school subjects, classes, and schools (i.e., to counteract the typical grading-on-a-curve effect). Furthermore, in both the original and expanded data sets comparable schools were selected on the basis of information supplied by the local Director of Education. Thus, grades were presumed to be comparable across schools (P. Huguet, personal communication, February 2, 2005). For this reason and because there was no other basis of scaling the class-average ability values for the different classes (e.g., the classes were not tracked in relation to student ability levels as in the Blanton et al., 1999 study), class-average grades were used as a basis for evaluating the BFLPE for the Huguet et al. (2001) data. However, in the absence of a standardised achievement test, this approach may have resulted in an underestimation of the BFLPE, due to the grading-on-a-curve effect. In fact, an examination of the ICCs for each subject in the expanded data is suggestive of a grading-on-a-curve effect. An inspection of the OECD PISA 2003 data (see Organisation for Economic Cooperation and Development, 2005a, 2005b) suggests that there are substantial differences between schools in France (ICC = .46, see Appendix B) when based on a nationally representative sample of schools and
standardised achievement tests. Nevertheless, in the expanded data there were minimal differences in achievement between schools (ICC for French = .01; for history/geography = .07; and for math = .09) and classes (ICC for French = .15; for history/geography = .05; and for math = .03; see Appendix B). That schools in France did differ in ability when standardised tests were used as seen in the PISA data, but that differences between schools and classes were minimal in the Huguet et al. (2001) expanded data suggests that teachers in the current study were indeed grading-on-a-curve. Consequently, results of the present study are likely to be conservative in terms of demonstrating the BFLPE.

Research Design

Both the original data on which Huguet et al.’s (2001) paper is based and the extended data were collected over the course of an academic school year. Grades were taken from school reports three times during the academic year, and students completed a questionnaire containing the comparison level choice and self-evaluation measures at the end of trimester 2.

Procedure and Statistical Analyses

Missing data. Although there were missing data for comparison choice, as was the case with the Blanton et al. (1999) data, the missing data were not imputed to ensure the correspondence between the original study and this further analysis.

Standardisation. In both the original data and the expanded data, grades and self-evaluation were standardised ($M = 0, SD = 1$) across the entire sample. Class-average ability scores were calculated by taking the average of students’ standardised grades at T2 in every class for each school subject. These class-average ability scores were not restandardised so that individual student grades and class average ability scores were measured on a common metric.

Replication of original results. Huguet et al.’s (2001) original results for grades and self-evaluation were replicated using regression analysis. The expanded data was also analysed using regression analyses to demonstrate that Huguet’s original results could be replicated in the larger data set. For both data sets, when T2 grade was the outcome variable, predictor variables were T1 grade, self-evaluation, and comparison choice 1 and 2. When self-evaluation was the outcome variable, T2 grade and
comparison choice 1 and 2 were entered as predictors. Both the original data and the expanded data were then analysed using multilevel modelling to ascertain whether the results of the regression analyses could be upheld using a stronger statistical method. The variables used in the multilevel modelling were those used in the regression analyses. All effects used in the multilevel modelling analyses, except the intercept, were fixed.

**BFLPE analyses.** The original data containing only 11 classes were really too small to provide an adequate test of the BFLPE (i.e., at the class level, the N is only 11). Consequently, to offer a more appropriate test of the BFLPE, the larger expanded data set was used in subsequent analyses. Hence, using the expanded data only, multilevel modelling analyses were conducted to test for the presence of the BFLPE. As with the Blanton et al. (1999) data, two sets of analyses were conducted. The first investigated the moderating effect of comparison choice on the BFLPE. For these analyses, the outcome variable was self-evaluation for each subject and predictors were T1 grades, class-average ability, comparison choice, and the interaction of class-average ability and comparison choice. The second set of analyses examined the moderating effect of individual ability on the BFLPE. In this instance the outcome variable was again self-evaluation for each subject and T1 grades, class-average ability, and the interaction of class-average ability and individual ability, as evidenced by T1 grades, were entered as predictors. As was the case for Blanton et al.’s (1999) data, there were three levels in these data: students at level 1, classes at level 2, and schools at level 3. None of the effects (with the exception of the intercept) was allowed to be random as it was not necessary to do so to test the hypotheses or answer the research questions posed.

As was the case for the Blanton et al. (1999) data, at higher levels in the Huguet et al. (2001) expanded data there was slight evidence of non-linearity and heteroscedasticity for some academic subjects. However, as the assumptions were not severely violated no action was taken.


**Preliminary Descriptive Statistics**

*Self-evaluation.* As seen in Table 5.14, consistent with the original Huguet et al. (2001) data, the scores for self-evaluation in the expanded data clustered around the midpoint of three for each academic subject.

<table>
<thead>
<tr>
<th>Subject</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>2.99</td>
<td>0.83</td>
</tr>
<tr>
<td>History/Geography</td>
<td>3.00</td>
<td>0.88</td>
</tr>
<tr>
<td>Math</td>
<td>2.90</td>
<td>1.02</td>
</tr>
</tbody>
</table>

*Grades.* Table 5.15 shows means and standard deviations for individual grades and choice 1 grades for the expanded data. Consistent with results from Huguet et al.’s (2001) original data, paired *t* tests (see Table 5.15), comparing individual grades and choice 1 grades, demonstrated that participants chose to compare with a better performing other in each subject. This difference was statistically significant for all three academic subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Individual Grade M</th>
<th>SD</th>
<th>Comparison Choice 1 Grade M</th>
<th>SD</th>
<th>SDD</th>
<th>df</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>11.38</td>
<td>2.88</td>
<td>12.13</td>
<td>2.93</td>
<td>2.78</td>
<td>1083</td>
<td>-8.57***</td>
</tr>
<tr>
<td>History/Geog</td>
<td>12.13</td>
<td>2.93</td>
<td>12.30</td>
<td>3.54</td>
<td>3.67</td>
<td>1062</td>
<td>-5.42***</td>
</tr>
<tr>
<td>Math</td>
<td>10.98</td>
<td>3.90</td>
<td>12.11</td>
<td>4.02</td>
<td>3.92</td>
<td>1077</td>
<td>-8.99***</td>
</tr>
</tbody>
</table>

*Note.* History/Geog = History/geography; SDD = Standard deviation of the difference. df and *t* values are from paired *t* tests which were used to determine if there were statistical differences between an individual’s grade and comparison choice’s grade. A tendency toward upward comparison is indicated by negative *t* values. ***p < .001
Results

Overview of Analyses

Six sets of analyses were conducted to test the hypotheses and research questions posed to satisfy the aims of Study 2b. Firstly, Huguet et al.’s (2001) original results for grades and self-evaluation were replicated using multiple regression (Hypotheses 1.1 and 1.2). Then, again using multiple regression, Huguet et al.’s results for grades and self-evaluation were demonstrated in an expanded sample, which was collected at the same time as the original data, but not previously reported (Hypotheses 2.1 and 2.2). For both sets of analyses, when the outcome variable was T2 grades, predictor variables were T1 grade, self-evaluation, and comparison choice grade. When the outcome variable was self-evaluation, predictors were T2 grade and comparison choice grade.

Subsequently, using multilevel modelling, Huguet et al.’s (2001) results for grades and self-evaluation were replicated in the original sample (Hypotheses 3.1 and 3.2), and in the expanded sample (Hypotheses 4.1 and 4.2). In these two sets of analyses, when the outcome variable was T2 grade, predictors were T1 grade, self-evaluation, and the comparison choice grade. When self-evaluation was the outcome variable, predictors were T2 grade, and the comparison choice grade.

The fifth set of analyses comprised multilevel modelling analyses using the expanded data to test for the presence of the BFLPE (Hypotheses 5.1 and 5.2) and to ascertain whether comparison choice moderated the BFLPE (Hypothesis 6.1 and Research Question 6.1). The outcome variable was self-evaluation for each subject, and predictors were T1 grades, class-average ability, comparison choice grades, and the interaction of class-average ability and comparison choice grades.

Lastly, multilevel modelling analyses were conducted using the expanded data to ascertain if the BFLPE was moderated by individual ability (Research Question 7.1). Self-evaluation for each subject was the outcome variable and predictors were T1 grades, class-average ability, and the interaction of class-average ability and individual ability (T1 grades).
Results for Hypothesis 1.1: Replication of Huguet et al.’s (2001) Results Utilising Multiple Regression – Grades

Hypothesis 1.1 predicted that the results of the Huguet et al. (2001) study would be replicated for grades whereby, based on multiple regression, for all seven school subjects T2 grade would be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice 1. Additionally, it was hypothesised that T2 grade in French, math, drawing, and technology, would be significantly positively predicted by T2 comparison choice 2.

Results for the replication of Huguet et al.’s (2001) data for grades using multiple regression analyses are shown in Table 5.16. For all seven school subjects T2 grade was significantly positively predicted by T1 grade (ranging from 0.39 for technology to 0.68 for history/geography), T2 self-evaluation (ranging from 0.11 for drawing to 0.35 for English), and T2 comparison choice 1 (ranging from 0.12 for math to 0.28 for technology). Additionally, T2 grades in French (0.11), math (0.12), drawing (0.28), and technology (0.19), were significantly positively predicted by T2 comparison choice 2, as per the original findings. Given that the original Huguet et al. results were replicated, Hypothesis 1.1 is accepted.

Results for Hypothesis 1.2: Replication of Huguet et al.’s (2001) Results Utilising Multiple Regression – Self-Evaluation

This hypothesis predicted that the results of the Huguet et al. (2001) study would be replicated for self-evaluation whereby, based on multiple regression, self-evaluation for all seven subjects would be significantly positively predicted by T2 grade but not by T2 comparison choice 1. Additionally it was predicted that T2 comparison choice 2 would only be significantly positively related to self-evaluation for biology; for all other academic subjects the effect of T2 comparison choice 2 would not be statistically significant.
Table 5.16. Recreation of Huguet et al.’s (2001) Original Data, Regressing T2 Grade on T1 Grade, T2 Self-Evaluation and T2 Comparison Choices 1 and 2, Using Standardised Regression Coefficients

<table>
<thead>
<tr>
<th>Subject</th>
<th>Huguet’s original results</th>
<th>Recreation</th>
<th>T1 Grade</th>
<th>T2 SE</th>
<th>T2 Choice 1</th>
<th>T2 Choice 2</th>
<th>df</th>
<th>T1 Grade</th>
<th>T2 SE</th>
<th>T2 Choice 1</th>
<th>T2 Choice 2</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td></td>
<td></td>
<td>0.56</td>
<td>0.23</td>
<td>0.20</td>
<td>-0.03</td>
<td>188</td>
<td>0.60</td>
<td>0.22</td>
<td>0.15</td>
<td>-0.02</td>
<td>195</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td>10.06***</td>
<td>4.36***</td>
<td>4.05***</td>
<td>-0.69</td>
<td></td>
<td>10.90***</td>
<td>4.24***</td>
<td>3.22**</td>
<td>-0.321</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
<td>0.61</td>
<td>0.16</td>
<td>0.17</td>
<td>0.10</td>
<td>216</td>
<td>0.64</td>
<td>0.16</td>
<td>0.15</td>
<td>0.11</td>
<td>223</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td>13.84***</td>
<td>4.14***</td>
<td>4.93***</td>
<td>3.35***</td>
<td></td>
<td>14.47***</td>
<td>4.17***</td>
<td>4.46***</td>
<td>3.44**</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td>0.50</td>
<td>0.34</td>
<td>0.17</td>
<td>0.06</td>
<td>200</td>
<td>0.51</td>
<td>0.35</td>
<td>0.17</td>
<td>0.06</td>
<td>209</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td>10.27***</td>
<td>7.28***</td>
<td>4.73***</td>
<td>1.71</td>
<td></td>
<td>10.89***</td>
<td>7.65***</td>
<td>4.81***</td>
<td>1.72</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History/Geography</td>
<td></td>
<td></td>
<td>0.68</td>
<td>0.19</td>
<td>0.17</td>
<td>-0.03</td>
<td>191</td>
<td>0.68</td>
<td>0.19</td>
<td>0.15</td>
<td>-0.02</td>
<td>200</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td>14.10***</td>
<td>4.07***</td>
<td>4.46***</td>
<td>-0.65</td>
<td></td>
<td>14.36***</td>
<td>4.41***</td>
<td>4.01***</td>
<td>-0.55</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td>0.64</td>
<td>0.17</td>
<td>0.15</td>
<td>0.11</td>
<td>176</td>
<td>0.66</td>
<td>0.15</td>
<td>0.12</td>
<td>0.12</td>
<td>190</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td>12.99***</td>
<td>3.38***</td>
<td>3.63***</td>
<td>2.92***</td>
<td></td>
<td>13.54***</td>
<td>3.18***</td>
<td>2.91**</td>
<td>3.02**</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing</td>
<td></td>
<td></td>
<td>0.42</td>
<td>0.10</td>
<td>0.30</td>
<td>0.28</td>
<td>171</td>
<td>0.40</td>
<td>0.11</td>
<td>0.25</td>
<td>0.28</td>
<td>181</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td>6.50***</td>
<td>1.72</td>
<td>4.20***</td>
<td>5.00***</td>
<td></td>
<td>6.87***</td>
<td>2.07*</td>
<td>4.34***</td>
<td>5.09***</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td></td>
<td></td>
<td>0.37</td>
<td>0.17</td>
<td>0.28</td>
<td>0.19</td>
<td>170</td>
<td>0.39</td>
<td>0.14</td>
<td>0.28</td>
<td>0.19</td>
<td>181</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td>5.34***</td>
<td>2.55***</td>
<td>4.51***</td>
<td>3.50***</td>
<td></td>
<td>6.22***</td>
<td>2.27*</td>
<td>5.00***</td>
<td>3.60***</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01; ***p < .001. Regression coefficients and degrees of freedom are slightly different in the re-creation. Huguet et al. included cross-products between comparison level choice and closeness, academic control, and importance of the academic domain. None of these interactions were significant and so were not added in the recreation.
Table 5.17 presents details of the replication of Huguet et al.’s (2001) data for self-evaluation, based on multiple regression. T2 grade was a statistically significant predictor of self-evaluation for all seven academic subjects tested, ranging from 0.39 for technology to 0.68 for history/geography. The relation between comparison choice 1 and self-evaluation was not statistically significant for any of the academic subjects tested. Comparison choice 2 was a statistically significant predictor of self-evaluation for biology only (0.16); for all other subjects comparison choice 2 was not significant. As the original Huguet et al. results were replicated, Hypothesis 1.2 is accepted.

**Summary for Hypotheses 1.1 and 1.2: Replication of Huguet et al.’s (2001) Results Utilising Multiple Regression – Grades and Self-Evaluation**

When T2 grade was the outcome variable, Huguet et al.’s (2001) findings were replicated using multiple regression. For each academic subject, previous ability, a positive evaluation of ability in that subject, and the comparison with a better performing target were predictive of higher grades at T2. The self-evaluation findings were also replicated. A positive evaluation of abilities in a subject was determined by ability in that particular subject. With the exception of comparison choice 2 for biology, comparisons with a selected individual had no impact on self-evaluation.

**Results for Hypothesis 2.1: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multiple Regression – Grades**

Hypothesis 2.1 predicted that the results of the Huguet et al. (2001) study would be replicated for grades in the expanded data, based on three school subjects (French, math, history/geography). It was hypothesised that, using multiple regression, for each school subject T2 grade would be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice 1. Additionally, it was predicted that T2 grades in French and math (but not history/geography) would be significantly positively predicted by T2 comparison choice 2.
Table 5.17. *Recreation of Huguet et al.’s (2001) Original Data, Regressing T2 Self-Evaluation on T2 Grade Comparison Choices 1 and 2, Using Standardised Regression Coefficients*

<table>
<thead>
<tr>
<th>Subject</th>
<th>T2 Grade</th>
<th>T2 Choice 1</th>
<th>T2 Choice 2</th>
<th>df</th>
<th>T2 Grade</th>
<th>T2 Choice 1</th>
<th>T2 Choice 2</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.55</td>
<td>-0.08</td>
<td>0.15</td>
<td>192</td>
<td>0.54</td>
<td>-0.02</td>
<td>0.16</td>
<td>197</td>
</tr>
<tr>
<td>t</td>
<td>8.47***</td>
<td>-1.20</td>
<td>2.62**</td>
<td></td>
<td>8.34***</td>
<td>-0.37</td>
<td>2.66**</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.82</td>
<td>-0.08</td>
<td>-0.06</td>
<td>222</td>
<td>0.80</td>
<td>-0.06</td>
<td>-0.07</td>
<td>226</td>
</tr>
<tr>
<td>t</td>
<td>12.59***</td>
<td>-1.28</td>
<td>-1.17</td>
<td></td>
<td>12.65***</td>
<td>-0.95</td>
<td>-1.21</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.81</td>
<td>-0.03</td>
<td>-0.06</td>
<td>207</td>
<td>0.81</td>
<td>-0.02</td>
<td>-0.05</td>
<td>212</td>
</tr>
<tr>
<td>t</td>
<td>15.40***</td>
<td>-0.62</td>
<td>-1.32</td>
<td></td>
<td>15.66***</td>
<td>-0.51</td>
<td>-1.06</td>
<td></td>
</tr>
<tr>
<td>History/Geography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.72</td>
<td>-0.09</td>
<td>-0.00</td>
<td>197</td>
<td>0.73</td>
<td>-0.09</td>
<td>0.09</td>
<td>203</td>
</tr>
<tr>
<td>t</td>
<td>12.22***</td>
<td>-1.55</td>
<td>1.55</td>
<td></td>
<td>12.40***</td>
<td>-1.45</td>
<td>1.63</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.62</td>
<td>-0.09</td>
<td>0.06</td>
<td>186</td>
<td>0.62</td>
<td>-0.07</td>
<td>0.08</td>
<td>193</td>
</tr>
<tr>
<td>t</td>
<td>9.55***</td>
<td>-1.40</td>
<td>0.94</td>
<td></td>
<td>9.24***</td>
<td>-1.06</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>Drawing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.36</td>
<td>0.10</td>
<td>-0.07</td>
<td>177</td>
<td>0.40</td>
<td>0.04</td>
<td>-0.09</td>
<td>184</td>
</tr>
<tr>
<td>t</td>
<td>4.12***</td>
<td>0.95</td>
<td>-0.86</td>
<td></td>
<td>4.42***</td>
<td>0.48</td>
<td>-1.08</td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.39</td>
<td>0.04</td>
<td>-0.00</td>
<td>177</td>
<td>0.41</td>
<td>0.05</td>
<td>-0.01</td>
<td>183</td>
</tr>
<tr>
<td>t</td>
<td>5.12***</td>
<td>0.59</td>
<td>-0.04</td>
<td></td>
<td>5.18***</td>
<td>0.66</td>
<td>-0.10</td>
<td></td>
</tr>
</tbody>
</table>

*Note.** *p < .01; ***p < .001. Regression coefficients and degrees of freedom are slightly different in the recreation. Huguet et al. included cross-products not incorporated in the recreation.*
Results of the replication of Huguet et al.’s (2001) original results for grades in the expanded data, using multiple regression, are presented in Table 5.18. In each school subject, T2 grade was significantly positively predicted by T1 grade (ranging from 0.49 for history/geography to 0.56 for French and math), T2 self-evaluation (ranging from 0.23 for French to 0.28 for math), and T2 comparison choice 1 (ranging from 0.13 for math to 0.14 for French and history/geography). However, not only did T2 comparison choice 2 significantly positively predict T2 grades in French (0.14) and math (0.09), but it was also a significant positive predictor of T2 grade in history/geography (0.12) in the expanded data.

Using multiple regression, Huguet et al.’s (2001) original results for grades were extended to Huguet et al.’s expanded data, with one minor exception: In Huguet et al.’s (2001) original data, T2 comparison choice 2 was not a significant predictor of T2 grades in history/geography, but was in the expanded data. For this reason, Hypothesis 2.1 is accepted for French and math, but not for history/geography.

Table 5.18. Multiple Regression Analysis of Huguet et al.’s (2001) Expanded Data, Regressing T2 Grade on T1 Grade, T2 Self-Evaluation and T2 Comparison Choices 1 and 2, Using Standardised Regression Coefficients

<table>
<thead>
<tr>
<th>Subject</th>
<th>T1 Grade</th>
<th>T2 SE</th>
<th>T2 Choice 1</th>
<th>T2 Choice 2</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>0.56</td>
<td>0.23</td>
<td>0.14</td>
<td>0.14</td>
<td>945</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>24.26*</td>
<td>10.69*</td>
<td>7.13*</td>
<td>7.28*</td>
<td></td>
</tr>
<tr>
<td>History/Geography</td>
<td>0.49</td>
<td>0.28</td>
<td>0.14</td>
<td>0.12</td>
<td>887</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>18.31*</td>
<td>10.98*</td>
<td>6.14*</td>
<td>5.32*</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>0.56</td>
<td>0.26</td>
<td>0.13</td>
<td>0.09</td>
<td>915</td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>22.74*</td>
<td>11.45*</td>
<td>6.54*</td>
<td>5.07*</td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p < .001

Results for Hypothesis 2.2: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multiple Regression – Self-Evaluation

Hypothesis 2.2 predicted that Huguet et al.’s (2001) results would be replicated for self-evaluation in the expanded data, based on three school subjects (French,
math, history/geography). Using multiple regression, it was predicted that for each school subject, self-evaluation would be significantly positively predicted by T2 grade, but T2 comparison choices 1 and 2 would not be significantly related to self-evaluation. Table 5.19 indicates that self-evaluation was positively predicted by T2 grade, ranging from 0.66 for history/geography to 0.73 for math. The relation between comparison choice 1 and self-evaluation was not significant for any of the academic subjects tested. Whereas comparison choice 2 was not significant for history/geography and math, there was a statistically significant, but small, negative relation between comparison choice 2 and self-evaluation for French (-0.07).

As regards self-evaluation, the expanded data demonstrated similar results to those of Huguet et al.’s (2001) original study, with one exception. Whereas comparison choice 1 did not predict self-evaluation for any subject, comparison choice 2 did, but for French alone. Students who compared with a better performing other had lower French self-concepts. Accordingly, Hypothesis 2.2 is accepted for math and history/geography, but not accepted for French.

Table 5.19. *Multiple Regression Analysis of Huguet et al.’s (2001) Expanded Data, Regressing T2 Self-Evaluation on T2 Grade and T2 Comparison Choices 1 and 2, Using Standardised Regression Coefficients*

<table>
<thead>
<tr>
<th>Subject</th>
<th>T2 Grade</th>
<th>T2 Choice 1</th>
<th>T2 Choice 2</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.68</td>
<td>0.01</td>
<td>-0.07</td>
<td>962</td>
</tr>
<tr>
<td>t</td>
<td>22.24**</td>
<td>0.31</td>
<td>-2.62*</td>
<td></td>
</tr>
<tr>
<td>History/Geography</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.66</td>
<td>-0.01</td>
<td>-0.03</td>
<td>905</td>
</tr>
<tr>
<td>t</td>
<td>22.27**</td>
<td>-0.49</td>
<td>-1.13</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Coefficient</td>
<td>0.73</td>
<td>0.00</td>
<td>-0.04</td>
<td>932</td>
</tr>
<tr>
<td>t</td>
<td>25.92**</td>
<td>0.02</td>
<td>-1.56</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .01; **p < .001
Summary for Hypotheses 2.1 and 2.2: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multiple Regression – Grades and Self-Evaluation

The original Huguet et al. (2001) results for grades and self-evaluation were replicated in the expanded data, with two exceptions. In the expanded data, T2 comparison choice 2 was a significant positive predictor of T2 grades for all three academic subjects and a significant negative predictor for self-evaluation of French. However, given the size of the sample, the effect of comparison choice 2 on French self-evaluation was trivial. Overall, however, these results demonstrated that higher grades in an academic course were influenced by individual ability, a positive self-evaluation in that course, and comparison with better performing others. Additionally, self-evaluation for students in the expanded data was associated with a student’s ability in that course and generally not with the performances of others.

Results for Hypothesis 3.1: Replication of Huguet et al.’s (2001) Results Utilising Multilevel Modelling – Grades

Hypothesis 3.1 predicted that Huguet et al.’s (2001) results would be replicated using multilevel modelling for all seven school subjects, whereby T2 grade would be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice 1. This hypothesis further predicted that T2 comparison choice 2 would be a significant positive predictor of T2 grades in French, math, drawing, and technology, but would not predict T2 grades in biology, English, and history/geography. Results of the replication of Huguet et al.’s (2001) findings for grades, based on multilevel modelling, are shown in Table 5.20. T2 grade was significantly positively predicted by T1 grade (ranging from 0.407 for technology to 0.693 for French), T2 self-evaluation (ranging from 0.111 for drawing to 0.224 for biology), and T2 comparison choice 1 (ranging from 0.101 for math to 0.294 for technology). T2 comparison choice 2 was a significant positive predictor of T2 grade for math, drawing, and technology, but not for French (see Table 5.20). T2 comparison choice 2 was not a significant predictor of self-evaluation for biology, French, English, and history/geography. Huguet et al.’s (2001) original findings were replicated using multilevel modelling, with one exception: T2 comparison choice 2
did not predict T2 French grade. Accordingly, Hypothesis 3.1 is accepted for all academic subjects tested except French.

**Results for Hypothesis 3.2: Replication of Huguet et al.’s (2001) Results Utilising Multilevel Modelling – Self-Evaluation**

Hypothesis 3.2 predicted that Huguet et al.’s (2001) self-evaluation results would be replicated for all seven school subjects when multilevel modelling was used. It was anticipated that self-evaluation would be significantly positively predicted by T2 grade, but comparison choice 1 would not be significant. It was further predicted that for French, English, history/geography, math, drawing, and technology, T2 comparison choice 2 would not be significantly related to self-evaluation, but that T2 comparison choice 2 would be a statistically significant positive predictor of self-evaluation for biology.

The results of the replication of Huguet et al.’s (2001) findings for self-evaluation using multilevel modelling are presented in Table 5.21. For all subjects, self-evaluation was significantly positively predicted by T2 grades ranging from 0.381 for drawing to 0.844 for French. T2 comparison choice 1 had no association with self-evaluation for any subject. T2 comparison choice 2 was not statistically significant for any subject except for biology where its effect was significantly positive (0.152).

Huguet et al.’s (2001) findings for self-evaluation were replicated using multilevel modelling. Self-evaluation in a subject was related to ability in that subject. Comparisons with a selected individual were not associated with self-evaluation, with the exception of comparison choice 2 for biology. In this instance, comparing with a better performing other was related to a higher self-evaluation in that subject. These results were consistent with those of the original study. Hence, Hypothesis 3.2 is accepted.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade (SE)</td>
<td>T2 Self-Eval (SE)</td>
</tr>
<tr>
<td>Biology</td>
<td>0.589* (0.053)</td>
<td>0.224* (0.052)</td>
</tr>
<tr>
<td>Drawing</td>
<td>0.435* (0.060)</td>
<td>0.111* (0.053)</td>
</tr>
<tr>
<td>English</td>
<td>0.582* (0.048)</td>
<td>0.304* (0.044)</td>
</tr>
<tr>
<td>French</td>
<td>0.693* (0.039)</td>
<td>0.164* (0.034)</td>
</tr>
<tr>
<td>History/Geography</td>
<td>0.688* (0.047)</td>
<td>0.194* (0.043)</td>
</tr>
<tr>
<td>Math</td>
<td>0.669* (0.047)</td>
<td>0.173* (0.046)</td>
</tr>
<tr>
<td>Technology</td>
<td>0.407* (0.064)</td>
<td>0.147* (0.064)</td>
</tr>
</tbody>
</table>

Note. All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. T2 Self-eval = T2 Self-evaluation. T2 Grade is the outcome variable. T1 Grade, T2 Self-evaluation, and T2 Choices 1 and 2 are predictors. Standard Error in parentheses (SE).

With the exception of the second comparison choice for French, Huguet et al.’s (2001) findings for grades were replicated using multilevel modelling. T2 grade in a particular subject was significantly positively predicted by ability in that subject, self-evaluation of ability, and by the first comparison choice. The second comparison choice was a significant positive predictor of T2 grades for three of the seven subjects tested. Huguet et al.’s results for self-evaluation were replicated using multilevel modelling. Self-evaluation in a particular subject was determined by ability in that subject and not by a student’s comparison choices, except for the second comparison choice in biology.


<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade</td>
<td>T2 Choice1</td>
</tr>
<tr>
<td>Biology</td>
<td>0.531*</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Drawing</td>
<td>0.381*</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.095)</td>
</tr>
<tr>
<td>English</td>
<td>0.811*</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>French</td>
<td>0.844*</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>History/</td>
<td>0.717*</td>
<td>-0.083</td>
</tr>
<tr>
<td>Geography</td>
<td>(0.057)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>Math</td>
<td>0.609*</td>
<td>-0.075</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Technology</td>
<td>0.386*</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.072)</td>
</tr>
</tbody>
</table>

Note. All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. T2 Self-eval = T2 Self-evaluation. T2 Self-evaluation is the outcome variable. T2 Grade, and T2 Choices 1 and 2 are predictors. Standard Error in Parentheses (SE).
Results for Hypothesis 4.1. Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multilevel Modelling – Grades

Hypothesis 4.1 predicted that Huguet et al.’s (2001) results for grades would be replicated for the expanded data using multilevel modelling based on three school subjects (French, math, history/geography). It was anticipated that for each school subject T2 grade would be significantly positively predicted by T1 grade, T2 self-evaluation, and T2 comparison choice 1. Additionally, it was anticipated that T2 grades in French and math (but not history/geography) would be significantly positively predicted by T2 comparison choice 2.

Table 5.22 presents the results of the replication of Huguet et al.’s (2001) findings for grades in the expanded data, using multilevel modelling. For each school subject, T2 grade was significantly positively predicted by T1 grade (ranging from 0.537 for history/geography to 0.599 for math), T2 self-evaluation (ranging from 0.218 for French to 0.258 for history/geography), and T2 comparison choice 1 (ranging from 0.094 for French to 0.98 for math). Contrary to the results for the original data, but consistent with the multiple regression results for the expanded data, not only did T2 comparison choice 2 significantly positively predict T2 grades in French (0.091) and math (0.070), but it was also a significant positive predictor of T2 grade in history/geography (0.064) in the expanded data.

Huguet et al.’s (2001) findings for grades were replicated in the expanded data using multilevel modelling, again with one difference as was seen in the multiple regression analyses of the expanded data: Not only did T2 comparison choice 2 significantly predict T2 grades in French and math, but it also significantly predicted T2 grades in history/geography. For this reason, Hypothesis 4.1 is accepted for French and math, but not accepted for history/geography.
Table 5.22. Multilevel Modelling Analyses of Huguet et al.’s (2001) Expanded Data, Predicting T2 Grades

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade (SE)</td>
<td>T2 Self-eval (SE)</td>
</tr>
<tr>
<td>French</td>
<td>0.597* (0.023)</td>
<td>0.218* (0.020)</td>
</tr>
<tr>
<td>History/Geography</td>
<td>0.537* (0.027)</td>
<td>0.258* (0.025)</td>
</tr>
<tr>
<td>Math</td>
<td>0.599* (0.026)</td>
<td>0.254* (0.022)</td>
</tr>
</tbody>
</table>

Note. All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. T2 Self-eval = T2 Self-evaluation. Ind = Individual. Time 2 Grade is the outcome variable. T1 Grade, T2 Self-evaluation, and T2 Comparison Choices 1 and 2 are predictors. Standard Error in parentheses (SE).
Results for Hypothesis 4.2: Replication of Huguet et al.’s (2001) Results Based on Expanded Data Utilising Multilevel Modelling – Self-Evaluation

This hypothesis anticipated that, using multilevel modelling, Huguet et al.’s (2001) results for self-evaluation would be replicated in the expanded data based on three school subjects. It was predicted that for each school subject self-evaluation would be significantly positively predicted by T2 grade, but T2 comparison choices 1 and 2 would not be statistically significant.

As indicated in the results presented in Table 5.23, self-evaluation for a particular subject was significantly positively predicted by a student’s T2 grade in that subject. Effects ranged from 0.670 for history/geography to 0.765 for math. The relation between comparison choice 1 and 2 and self-evaluation was not statistically significant for any of the three subjects tested. As Huguet et al.’s (2001) results for self-evaluation were replicated in the expanded data using multilevel modelling, Hypothesis 4.2 is accepted.


Using multilevel modelling, Huguet et al.’s (2001) findings for grades were replicated in the expanded data, with the same exception seen in the multiple regression analyses regarding T2 comparison choice. These findings offer evidence that, comparing upward, having a positive self-evaluation of one’s academic achievements, and initial ability, all have a beneficial impact on subsequent grades.

Huguet et al.’s (2001) findings were replicated for self-evaluation. Although multiple regression analyses of the expanded data had revealed that comparison choice 2 had a statistically significant effect on self-evaluation for French, the size of the effect was trivial. This result was not replicated using the more stringent statistical method of multilevel modelling. Ability in a particular subject was the primary predictor of self-evaluation and comparisons with selected others were not related to self-evaluations of ability in that subject.
Table 5.23. Multilevel Modelling Analyses of Huguet et al.’s (2001) Expanded Data, Predicting Self-Evaluation

<table>
<thead>
<tr>
<th>Subject</th>
<th>T2 Grade (SE)</th>
<th>T2 Choice 1 (SE)</th>
<th>T2 Choice 2 (SE)</th>
<th>Constant (SE)</th>
<th>School (SE)</th>
<th>Class (SE)</th>
<th>Ind (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>0.720*</td>
<td>0.035</td>
<td>-0.029</td>
<td>0.042</td>
<td>0.006</td>
<td>0.057*</td>
<td>0.509*</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.028)</td>
<td>(0.046)</td>
<td>(0.013)</td>
<td>(0.020)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>History/Geography</td>
<td>0.670*</td>
<td>0.012</td>
<td>-0.002</td>
<td>0.059</td>
<td>0.019</td>
<td>0.016</td>
<td>0.557*</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.030)</td>
<td>(0.028)</td>
<td>(0.046)</td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Math</td>
<td>0.765*</td>
<td>0.024</td>
<td>-0.020</td>
<td>0.055</td>
<td>0.018</td>
<td>0.016</td>
<td>0.452*</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.027)</td>
<td>(0.025)</td>
<td>(0.044)</td>
<td>(0.012)</td>
<td>(0.010)</td>
<td>(0.021)</td>
</tr>
</tbody>
</table>

*Note. All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. T2 Self-eval = T2 Self-evaluation. Ind = Individual. T2 Self-evaluation is the outcome variable. Time 2 Grade and T2 Comparison Choices 1 and 2 are predictors. Standard Error in parentheses (SE).
**Results for Hypothesis 5.1: Tests of the BFLPE in the Huguet et al.’s Expanded Data Based on Multilevel Modelling – The Relation Between Individual Ability and Self-Evaluation**

Hypothesis 5.1 predicted that, using multilevel modelling, for each school subject in the Huguet et al. expanded data, self-evaluation would be significantly positively predicted by individual ability as measured by T1 grade. Results indicated that for all three school subjects of the Huguet et al. expanded data, the relation between individual ability and self-evaluation was statistically positively significant, ranging from .641 for French to .745 for math (see Table 5.24). As there was a significantly positive relation between individual ability and self-evaluation in the Huguet et al. (2001) expanded data Hypothesis 5.1 is supported.

**Results for Hypothesis 5.2: Tests of the BFLPE in Huguet et al.’s (2001) Expanded Data Based on Multilevel Modelling – The Relation Between Class-Average Ability and Self-Evaluation**

Hypothesis 5.2 predicted that, using multilevel modelling, for each school subject in the Huguet et al. expanded data, self-evaluation would be significantly negatively predicted by class-average ability, as measured by T2 class-average grade. Results of analyses demonstrated that the relation between class-average ability and self-evaluation was negative for all three subjects (see Table 5.24), with the effect being statistically significant in French (-.330) and math (-.415), but not for history/geography. Hence, the hypothesis that there would be a significantly negative relation between class-average ability and self-evaluation was supported for two of the three academic subjects tested. Consequently, Hypothesis 5.2 is accepted for French and math, but not for history/geography.

**Summary for Hypotheses 5.1 and 5.2: Tests of the BFLPE in Huguet et al.’s (2001) Expanded Data Based on Multilevel Modelling**

Results for the Huguet et al. expanded data demonstrated a significantly positive relation between individual ability and self-evaluation for all academic subjects, and a significantly negative relation between class-average ability and self-evaluation for two of the three academic subjects tested. These results signify that a BFLPE was present for French and math in the Huguet et al. expanded data.

This hypothesis anticipated that, for each school subject, the comparison person’s grade would not be statistically significantly related to self-evaluation in the Huguet et al. expanded data. However, results based on multilevel modelling demonstrated that the comparison person’s grade had a small positive association with self-evaluation that varied from .04 to .12. This effect was statistically significant for French (.115) and math (.074), but not significant for history/geography (see Table 5.24). Given that the comparison person’s grade had a statistically significant association with self-evaluation for two of the three subjects tested, Hypothesis 6.1 was not accepted for French and math, but was accepted for history/geography.


Research Question 6.1 enquired whether, for each school subject, the comparison person’s ability could moderate the effect of class-average ability on self-evaluation in the Huguet et al. expanded data, or whether the BFLPE co-existed with the positive effects of upward comparison on performance. Multilevel modelling analyses of the Huguet et al. expanded data indicated that the class-average ability X comparison choice interaction was not statistically significantly related to self-evaluation for any of the three school subjects tested (see Table 5.24).

Summary for Hypothesis 6.1 and Research Question 6.1: Comparison Choice

In the Huguet et al. expanded data those who chose more able comparison targets had significantly higher self-perceptions of their academic ability in French and math, but not for history/geography. These results for French and math are contrary to Huguet et al.’s (2001) original results, based on a smaller sample, in which comparison choice 1 had no effect on self-evaluation. Additionally, analyses of the expanded data demonstrated that the comparison person’s grade did not moderate the effect of class-average ability on self-evaluation (i.e., the BFLPE) for all three academic subjects tested. Thus, no matter with whom students chose to compare, the higher the ability of the class, the lower their self-evaluations.
Table 5.24. *Multilevel Modelling Analyses of Huguet et al.’s (2001) Expanded Data Predicting Self-Evaluation, with Comparison Choice as a Moderator of the BFLPE*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade</td>
<td>School</td>
</tr>
<tr>
<td></td>
<td>T2 Class-average</td>
<td>Class</td>
</tr>
<tr>
<td></td>
<td>T2 Choice</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2 Choice X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2 Class-average</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>0.641*</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.015)</td>
</tr>
<tr>
<td></td>
<td>-0.330*</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(0.010)</td>
</tr>
<tr>
<td></td>
<td>0.115*</td>
<td>0.541*</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.024)</td>
</tr>
<tr>
<td></td>
<td>-0.064</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td></td>
</tr>
<tr>
<td>History/Geography</td>
<td>0.643*</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.016)</td>
</tr>
<tr>
<td></td>
<td>-0.122</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.010)</td>
</tr>
<tr>
<td></td>
<td>0.041</td>
<td>0.582*</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.026)</td>
</tr>
<tr>
<td></td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>0.745*</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.008)</td>
</tr>
<tr>
<td></td>
<td>-0.415*</td>
<td>0.024*</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.011)</td>
</tr>
<tr>
<td></td>
<td>0.074*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.066</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). Regression coefficients are based on standardised scores. Standard Error in Parentheses (SE).
Results for Research Question 7.1: Individual Ability as a Moderator of the BFLPE in Huguet et al.’s (2001) Expanded Data Based on Multilevel Modelling

Research Question 7.1 enquired about the nature of the relation between the individual ability X class-average ability interaction and self-evaluation for each school subject in the Huguet et al. expanded data. When individual ability was considered as a moderator of the BFLPE in the Huguet et al. expanded data, a BFLPE emerged for French and math, but not for history/geography (see Table 5.25). Individual ability was a significant positive predictor of self-evaluation for all three school subjects (ranging from .699 for French to .765 for math), but class-average ability was a significant negative predictor for only French and math (-.261 and -.352 respectively). Results for the moderating effect of individual ability signified that whereas the ability-by-class-average interaction was not significant for history/geography and French, the interaction was statistically significantly negative for math (-.164). Figure 5.3 demonstrates the pattern of this interaction showing that the size of the BFLPE was larger for more able students. Compared to students of similar ability in low-achieving classes, students of low and average ability had slightly lower self-evaluations if they attended high-ability classes. A larger drop in self-evaluation was seen for students of high-ability if they attended high-ability classes. In this case, compared to students of similar ability in low-achieving classes, high-ability students had considerably lower self-evaluations if they attended high-ability classes.

Study 2b Summary

Study 2b, a further analysis of the Huguet et al. (2001) data, was described in this section. As was the case for Study 2a, its main goal was to determine if selected comparisons moderated, or co-existed with, the BFLPE. The section provided specific aims, hypotheses and research questions, rationales for the hypotheses and research questions, methodology, and results for Study 2b. Using multiple regression, Huguet et al.’s (2001) original results for grades and self-evaluation were replicated. With two exceptions, both for the second comparison choice, they were also replicated for the expanded data. Using multilevel modelling, results for self-evaluation were replicated for both the original and expanded data. Multilevel
modelling analyses for grades indicated that the original results were replicated for the original and expanded data, with exceptions again for the second comparison choice. Analyses also indicated that a BFLPE was present for two of the three subjects tested in the expanded data, but that the BFLPE was not moderated by comparison choice. However, individual ability moderated the BFLPE for one subject, whereby larger BFLPEs were associated with higher ability students in math. The chapter now proceeds with a discussion of the findings of Studies 2a and 2b.

Figure 5.3. Math Ability by Class-Average Ability Interaction in Huguet et al.’s (2001) Expanded Data

Note. Based on predicted values. High class-average ability = 1 standard deviation above the mean for class-average math ability and low class-average ability = 1 standard deviation below the mean. Similarly, high-ability = 1 standard deviation above the mean for math ability, and low-ability = 1 standard deviation below the mean.
Table 5.25. *Multilevel Modelling Analyses of Huguet et al.’s (2001) Expanded Data Predicting Self-Evaluation, with Individual Ability as a Moderator of the BFLPE*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 Grade</td>
<td>T2 Class-average Ability</td>
</tr>
<tr>
<td>French</td>
<td>0.699* (0.024)</td>
<td>-0.261* (0.077)</td>
</tr>
<tr>
<td>History/ Geography</td>
<td>0.674* (0.025)</td>
<td>-0.126 (0.101)</td>
</tr>
<tr>
<td>Math</td>
<td>0.765* (0.024)</td>
<td>-0.353* (0.085)</td>
</tr>
</tbody>
</table>

*Note.* All parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two Standard Errors (SEs). Regression coefficients are based on standardised scores. Standard Error in Parentheses (SE).
Discussion

Overview

The current study was conducted to reconcile results from two strands of research that had produced conflicting results. The social comparison studies of Blanton et al. (1999) and Huguet et al. (2001) demonstrated that upward comparisons were associated with better academic performance. Conversely, BFLPE theory and research contends that social comparisons are responsible for lowered academic self-concepts, and in turn, lower academic self-concepts have been associated with poorer academic performance (e.g., Marsh & Yeung, 1997a); see Chapters 2 and 3. Therefore, the principal aim of the current study was to ascertain whether selected individual comparisons could moderate the BFLPE, or whether the negative effects of the BFLPE co-existed with the beneficial effects of upward comparisons on performance. In reconciling these opposing findings, Study 2 began by replicating the results of the social comparison studies and continued by investigating whether they contained evidence of a BFLPE. Results supported the co-existence hypothesis, as the BFLPE was moderated by comparison choice in only one academic subject out of the 10 tested. Similarly, the results of analyses investigating if the BFLPE varied as a function of individual ability showed that individual ability moderated the BFLPE for only one academic subject. These results are discussed in more detail in the following sections.

Replications

The original Blanton et al. (1999) and Huguet et al. (2001) findings were replicated for grades and self-evaluation using multiple regression. Moreover, using the stronger statistical method of multilevel modelling, upward comparisons still had a positive effect on subsequent achievement, as was shown in the original studies, and self-evaluation in a particular subject was associated with ability in that subject, not with comparisons with a selected other.

Huguet et al.’s (2001) original results for grades and self-evaluation were also replicated in their expanded data, using multiple regression and multilevel modelling. However, there were two minor exceptions. In the original analyses, based on 264
students, history/geography grades were not predicted by the second comparison choice, but in the expanded data, based on 1,156 students, they were, and this was the case for both the multiple regression and the multilevel modelling analyses. Additionally, when the expanded data were analysed using multiple regression, self-evaluation was negatively predicted by the second comparison choice for French. These differences could be attributable to the larger sample size providing greater power. Nevertheless, it could also be that because students in the expanded data were in their first to fifth year of high school, those who were older may have had more experience with their second comparison choice.

Interestingly, when multilevel modelling was used to analyse Huguet et al.’s expanded data, the significant negative relation between the second comparison choice and French self-evaluation, found using multiple regression, was no longer significant. This was also the case for French grades in the original sample. Using multiple regression, French grades were significantly predicted by the second comparison choice, although the effect was small. Using multilevel analyses, this effect was no longer significant. Single level multiple regression ignores the hierarchical structure of these data, where students are nested within classes within schools. Traditional single level techniques that ignore such a multilevel structure are likely to be invalid as they not only violate assumptions of independence, but they also increase the likelihood of finding statistical significance where none exists (Hox, 2002), and this could be the case here. When multilevel modelling analyses were used to analyse these data, the variance associated with different effects would have been partitioned into components associated with the different levels, with the result that the small effects were no longer statistically significant. This finding, that significant results using multiple regression were no longer significant when using multilevel modelling, highlights the need to use more appropriate statistical techniques when data contain more than one level.

**Comparison Choice**

With one exception (biology for comparison choice 2 in the Huguet et al. (2001) study), the original Blanton et al. (1999) and Huguet et al. (2001) studies found that comparing with someone performing better than oneself was not related to self-
evaluations of ability. However, results from the BFLPE analyses of the Huguet et al. expanded data demonstrated that when students compared with someone performing better than they themselves were, their self-evaluations increased for two of the three subjects examined (French and math). The positive effects on self-evaluation of choosing a more able comparison target found in the expanded data may be due to students identifying with their comparison targets. In the Huguet et al. (2001) study, comparison choice predicted a participant’s tendency to regard the position of close comparison others as his/her own future. On this basis, the authors suggested that students had identified with their more successful and close comparison targets. The positive effects of comparison choice on self-evaluation, found in the expanded data, could simply be the consequence of this identification. However, since results across the two data sets were not consistent, no definitive conclusions can be drawn.

Interestingly, the positive associations between comparison choice and self-evaluation were not evident in the replication of Huguet et al.’s original results, but were when the BFLPE was examined (i.e., when class-average ability was added to the model). This suggests that comparison choice could consist of two components: one positive and one negative that cancel each other out. When the negative component of the comparison is controlled, as is the case when class-average ability is included (the BFLPE), the positive component becomes significant, thus suggesting that the BFLPE is indeed based on social comparisons. However, as this result was evident in only 2 of the 10 academic subjects tested, further research is needed to test this proposition.

**The BFLPE**

Although not designed to study the BFLPE, further analyses of the Blanton et al. (1999) data and analyses of the Huguet et al. (2001) expanded data provided evidence of a BFLPE in both the Netherlands and France, in the same data that demonstrated the beneficial effects of upward comparisons on performance. These results are consistent with previous research which has found that being in high-ability schools or classes has a negative impact on academic self-concept (e.g., Craven et al., 2000; Marsh, 2005; Marsh & Craven, 2002; Marsh & Hau, 2003; Marsh et al., 2001; Marsh & Rowe, 1996; Marsh et al., in press; Mulkey et al., 2005).
The findings from the current study are also consistent with those of the majority of social comparison research, which has shown that upward comparisons typically have negative effects on self-evaluations (e.g., Collins 1996, 2000; Lockwood & Kunda, 1997; Suls & Wheeler, 2000). Students in high-ability classes (i.e., classes in which the average ability level of students was high) would have been more exposed to high-achieving same-age peers than equally able students in low-ability classes, and this may have left them feeling that they were less adequate in comparison.

The major focus of the current investigation was whether the BFLPE co-existed with, or was moderated by, selected individual comparisons. In only one instance (Dutch) the BFLPE was moderated by selected comparisons; in all other cases (9 out of 10 tests) the BFLPE was not moderated by comparisons chosen by students. Consequently, these findings demonstrated that lower self-evaluations associated with being in a high-ability class (the BFLPE) could co-exist with – and were not moderated by – the upward social comparisons with individual students that enhanced performance. BFLPE theory suggests that social comparisons are responsible for the effect. In support of this proposition, these results suggest that the BFLPE may not be the result of upward comparisons in which students engage spontaneously, but rather, it may be the result of a comparison forced on the individual by the environment. In this, these results support the distinction made by various researchers (Diener & Fujita, 1997; Goethals, 1986; Suls, 1986) between forced and selective approaches to social comparison. In the latter, individuals choose their comparison target, but in the forced approach, “comparisons are imposed or compelled” (Diener & Fujita, 1997, p. 330) on the individual by the environment in which they find themselves. Whereas the effects of choosing a more able comparison target (representative of the selective approach) were never negative and were positive in some instances, the effects of the forced comparison (class-average ability) on self-evaluation were consistently negative (and significantly so in all but one instance) in the Blanton et al. (1999) and the expanded Huguet et al. (2001) data. Thus, the current study demonstrates that these two social comparison effects, selective and forced, can exist side by side, and simultaneously produce different outcomes. Apparently, the effect of choosing a specific comparison target differs from that of a forced evaluation of oneself against a group of other students.
Additionally, previous social comparison research has suggested that the selection of individual comparison targets is strategically motivated by self-evaluative, self-improvement, or self-enhancement goals (for a review, see Wood, 1989). However, it is unlikely that such motivations exist in the forced comparison paradigm typical of BFLPE studies because comparison targets are externally imposed. It is not surprising, perhaps, that results of selected comparisons differ from those that are externally imposed. Particularly when students are explicitly asked to choose the specific students with whom they are comparing themselves, it is not surprising that the comparison processes and integration of this information into the formation of the self-concept differ from those that occur when students evaluate themselves in relation to a broader, more diverse group of other students.

The Selective Accessibility Model

That the negative effects of the BFLPE on self-evaluation can co-exist with the positive effects on performance of upward social comparisons may also be consistent with the Selective Accessibility (SA) model (Mussweiler, 2001b; Mussweiler & Strack, 2000a). This model suggests that when making social comparisons, people either test for similarities or dissimilarities. So, if people look for similarities when making a comparison, this should result in assimilation and is probably the most common choice. When people look for dissimilarity, contrast should be the result. Mussweiler also assumed that regardless of whether people look for similarities or differences, because they have vast amounts of self-knowledge, they would find information consistent with whichever hypothesis they were testing. Empirical support for this model has been demonstrated in various studies (Mussweiler, 2001b; Mussweiler & Strack, 1999, 2000b).

Relating the current investigation to this model, perhaps when choosing a comparison target students chose one whose success was considered to be attainable. As Huguet et al. (2001) noted, students identified with their chosen comparison targets, reporting that their performance could become closer to that of their comparison choice in the future. Additionally, students in the current study chose comparison targets who did not harm their self-image, as can be seen by the null or very small positive effect of comparison choice on self-evaluation. Thus, comparison
targets may have been chosen because their success was attainable and comparisons with them were not self-deflating. When making their chosen comparisons, our students may have looked for, and found, similarities between themselves and their targets, thus leading them to assimilate their performances with their targets. The fact that selected upward comparisons did not lower self evaluations, and that students making use of this comparison see their more successful targets as possible future selves (assimilation), both suggest that upward comparison may be used as a means of self-improvement (see Wood, 1996) rather than as a means of self-enhancement. Conversely, forced comparisons, like those with the class-average, do not allow this kind of choice. So, students may have had to consider those who were definitely superior and whose success was unattainable, thus acknowledging differences between themselves and their targets. Accordingly, when evaluating their abilities, our students may have looked for, and found, differences between themselves and their targets, thus leading them to contrast their self-evaluations away from these targets. These speculations should be the focus of future research.

**Individual Ability**

In the current study the negative effect of the BFLPE generalised across all ability levels for seven school subjects in the Blanton et al. (1999) data and for two of three subjects in the Huguet et al. expanded data. However, there was a significant interaction effect for math in the Huguet et al. expanded data, such that, compared to lower-achieving students in high-ability classes, larger BFLPEs were associated with higher-achieving students. Nonetheless, the size of the statistically significant interaction was small and most of the interactions (9 of 10 tests) were not statistically significant. Although this interaction has been assessed in previous studies (Coleman & Fults, 1985; Marsh et al., 1995; Marsh & Hau, 2003; Marsh & Rowe, 1996; Reuman, 1989), results have shown it to be small and inconsistent in direction and in statistical significance. As a result, pending further research regarding the circumstances under which individual ability does/does not moderate the BFLPE, it must be concluded that the BFLPE is robust and generalises reasonably well over students of different ability levels.
Strengths and Limitations

Bringing together two theoretical perspectives in the context of a single study is a major strength of this study. Whereas social comparison research has demonstrated the beneficial effects of selected upward comparisons on performance, BFLPE research has proposed that comparisons have a deleterious effect on academic self-concept, and lower academic self-concept has been associated with lower achievement. To date, few studies have endeavoured to bring these two strands of research together. Hence, in doing so, the current study attempts to demonstrate the different effects associated with selected upward comparisons and forced comparisons with a generalised other.

Re-analysis of previously published data using stronger statistical methodology is another important contribution of Study 2. Although the majority of results were replicated, there were two anomalies: Previously statistically significant results were no longer significant when multilevel modelling was utilised. This finding underscores the need to use statistical methods that suit the data. Especially when the data contain different levels nested within each other, as was the case with these data, statistical methods that take these levels into account are essential.

However, interpretations of the current study must be made cautiously as, apart from the usual caveats that apply to correlational data, there are other potentially important limitations that warrant consideration. In particular, neither data set met the criteria of an ideal BFLPE study – which is not surprising, given that this was not their original intention. Neither the Blanton et al. (1999) study, nor the Huguet et al. (2001) expanded data included a standardised academic achievement test or a measure of academic self-concept, both of which are routinely used to test for the BFLPE, and so it is not unexpected that there are limitations in the current study in terms of these tests. In the present investigation, school grades, with all their inherent inconsistencies, were used as a performance measure, and a single item self-evaluation measure was used to assess academic self-concept. The reliance on a single self-evaluation item instead of a multi-item academic self-concept scale, with good psychometric properties, is a potentialshortcoming. However, estimates of the BFLPE based on this single-item self-evaluation measure were comparable with estimates based on the nationally representative PISA data for these countries as
demonstrated in Study 1 of the current thesis (effect of school-average ability on math self-concept for the Netherlands = -.696, and for France = -.383; see Table 4.7). As argued previously, the use of school grades instead of achievement, that would have allowed class-average achievement to be scaled along a metric that was common to all classes, is likely to be conservative. Furthermore, even though this issue was dealt with in different ways for the Blanton et al. (1999) and Huguet et al. (2001) data (see discussion below), the results of the re-analyses of the two studies were consistent with each other.

The measures of academic achievement used in this study were also problematic. For the Blanton et al. (1999) data, it was possible to scale classes in relation to ability tracks assigned to individual students by schools. While it was not possible to control the assignment of students to schools, it can be reasonably assumed that, on average, students placed in higher tracks were brighter than students placed in lower tracks. Nevertheless, it could be argued that the scale used to modify the grades was somewhat arbitrary and even conservative. Compared to the actual differences between schools observed in the PISA 2003 data (see Organisation for Economic Cooperation and Development, 2005a, 2005b), the scale used to transform the Blanton et al. data was conservative. Thus, it is probable that the size of the BFLPE evident in the Blanton et al. (1999) re-analysis is conservative compared to that which would have been found had a more suitable measure of academic achievement been used. The size of the BFLPE may also have been underestimated in the Huguet et al. (2001) expanded data. In that study, although comparable schools were selected, it is probable that the teachers, like teachers all over the world, graded on a curve at least to some extent. If so, then the differences observed in class-average grades would have greatly underestimated the actual differences in class-average grades, and subsequently the BFLPE (the negative effect of class-average achievement) would also have been underestimated. While the limitations of the French and Dutch studies are different, it is important to note that both are apparently conservative in their estimation of the BFLPE. As such, that such strong support was found for the BFLPE is remarkable and testifies to the robustness of the effect.
Implications for Theory, Research, and Practice

By demonstrating that selected and forced comparisons can exist side by side and simultaneously yield differing results, the current study has implications for BFLPE theory and for social comparison theory. Findings from the current study suggest that the BFLPE may not be based on comparisons that individuals select. Perhaps the BFLPE may be the result of comparisons with a generalised other, forced by the environment, or a combination of selected individual comparisons and forced generalised comparisons. Although social comparison theory and research acknowledge the existence of comparisons forced by the environment (e.g., Festinger, 1954; Wood, 1989) and that comparisons can be made with a generalised other (e.g., Festinger, 1954; Klein, 1997), very little is known about either of these types of comparisons (see Chapter 3). For example, Wood (1989) noted that social comparison research has for too long emphasised the role of target selection. As such, more prominence needs to be placed on comparisons imposed by the environment over which the individual has no choice. Additionally, when discussing specific and generalised comparisons, Buckingham and Alicke (2002) stated, “Social comparison theories have yet to consider the relative importance of these two data sources for establishing enduring conceptions of one’s abilities and characteristics” (p. 1128). Hence, to inform theory, research is needed to explicate the form and nature of the generalised other in regards to the BFLPE and to elucidate the mechanisms of forced comparisons.

The current study highlights several limitations in existing research, which future research could address. For example, the use of different methodologies across studies may be one cause of conflicting results. In some studies, such as the Blanton et al. (1999) and the Huguet et al. (2001) studies, students are explicitly asked to select comparison targets. Conversely, the common method used to examine the BFLPE represents a forced comparison paradigm (Diener & Fujita, 1997) where comparisons with the group-average are implicit. Explicit and implicit social comparisons have been shown to produce different outcomes. For example, Stapel and Suls (2004) noted that implicit comparisons result in contrast, but that under certain conditions explicit comparisons can produce assimilation. This distinction is apparently relevant in that comparison with a generalised other or class-average as implied in BFLPE research is implicit, whereas much of social comparison theory is
based on explicit comparisons. Future research could use the same methodology (i.e., either explicit selections, or forced implicit comparisons) within the same study to examine the effects of comparisons on performance and self-concept. Additionally, future research could determine whether the explicit selection of comparison targets and forced implicit comparisons constitute the same or different processes.

Furthermore, it could be argued that pre-existing individual differences that lead a person to choose someone as a comparison target may be more important in determining performance than the actual comparison. For example, students high in mastery motivation orientation may feel confident enough in their abilities to choose a comparison target who is currently performing slightly better, because they feel capable of performing just as well as the comparison target, or may be able to use the information gained from the comparison to self-improve. Individual differences such as these could be examined to ascertain whether these individual differences and not the comparison per se are the key to improved performance.

As previously noted by Marsh and colleagues, there is growing evidence that being in a school or class with high-achieving students is likely to undermine academic self-concept, that these effects are long lasting, and that they have negative implications for future academic accomplishments. This should be a cause for concern for parents and policy makers, since the development of a positive academic self-concept is an important educational goal and a determinant of future academic achievement. Hence, it would seem timely for policy makers to address, by means of methodologically sound state-of-the-art research, the negative consequences of segregating students on the basis of their academic abilities. For example, expanding the selection processes to these classes and schools to include criteria other than academic achievement may prove beneficial, as may interventions that target enhancing self-concept. Additionally, encouraging these selective environments to be more cooperative and less competitive may help to reduce the social comparison processes thought to underlie the BFLPE (see Marsh et al., 1995).

Interestingly, the present findings offer first evidence to date that the BFLPE co-exists with a positive effect of upward comparison choices on grades and that students in both high- and low-ability schools make these comparison choices. Evidently, most students choose to engage in (slightly) upward comparisons that
under appropriate conditions can have a beneficial effect on their performance. Therefore integrating students in less academically selective schools does not mean that they will not be able to engage in these potentially beneficial comparisons. At the practical level, integrating students in academically heterogeneous schools may simultaneously protect them from the BFLPE and allow them to benefit from selected upward comparisons.

Given these findings it is recommended that the BFLPE and comparison processes be examined simultaneously, using both academic self-concept and standardised performance measures. Researchers should be encouraged to pursue multi-method studies that combine quantitative approaches, like those used here and in other BFLPE studies, with qualitative approaches (e.g., interviews or diaries) that might provide better insights into the actual psychological processes underlying the BFLPE (e.g., Tracey et al., 2003).

In summary, the current study contributes to the idea that comparisons can be a “double-edged sword” (Major et al., 1991, p. 238; see also Buunk et al., 1990; Diener & Fujita, 1997; Taylor et al., 1990; Testa & Major, 1990). Apparently, comparisons can enhance performance, but at the same time they can make one feel distinctly inferior about one’s abilities. Choosing to compare with more able individual students might lead to improved academic performance in appropriate circumstances, but being in a class with high-achieving students is likely to undermine academic self-concept.

Summary

Study 2 of the present investigation was described in the current chapter. Firstly, the discrepancy between social comparison and BFLPE results was discussed and the collaborative background to Study 2 described. Then, further re-analyses of the two social comparison studies, on which Study 2 was based, were presented. The aims, hypotheses, research questions, rationales for hypotheses and research questions, methodology, and results were presented separately for each re-analysis. In essence, the results of Study 2 demonstrated that the BFLPE co-existed with, but was not moderated by, upward social comparisons that enhanced performance. Additionally,
it was concluded that the BFLPE generalised reasonably well over students of
different ability levels. The chapter concluded by discussing the results of both re-
analyses together. Strengths and limitations of Study 2, as well as implications for
theory, practice, and future research, were also discussed.
CHAPTER 6

STUDY 3: SELECTED SOCIAL COMPARISONS AND THE BFLPE

Introduction

Study 2 demonstrated that the negative effects of the BFLPE co-existed with the positive effects of upward social comparisons, and social comparisons have been implicated as one of the principal causes of the BFLPE. However, as the data analysed in Study 2 were originally collected to study the effect of social comparison on performance and not the BFLPE, it was not an ideal tool to elucidate the relation between the two processes. Additionally, although the theoretical model underpinning the BFLPE has proposed that it is the comparisons that students make with their peers that contribute to the BFLPE (e.g., Marsh & Hau, 2003), to date very little research has primarily investigated social comparison processes in the context of the BFLPE. Study 3 sought to rectify these limitations by examining the social comparisons students select and the extent of the BFLPE together in a single study. In doing so, Study 3 was able to overcome the main limitations of Study 2, these being the lack of a standardised test to measure ability and the lack of a psychometrically sound measure of academic self-concept. Study 3 also analysed potential moderators of the BFLPE to ascertain whether any of these could moderate its negative effects, thereby extending knowledge of the BFLPE.

The current chapter aims to describe Study 3 by detailing: (a) the problem being studied; (b) the background to the problem; (c) the general aims; (d) the hypotheses and research question under investigation; (e) the rationale for hypotheses and research questions; (f) the methodology used; and (g) results of the investigation in the context of the hypotheses and research questions posed. The chapter concludes with a discussion of the results and the implications thereof for theory, research, and practice.
Statement of the Problem

What is the relation between the BFLPE and social comparisons that individuals select? Does the BFLPE vary as a function of individual ability? Can any factors be identified that moderate the negative effects of the BFLPE?

Background

As a result of the joint venture undertaken in Study 2, Pascal Huguet, Florence Dumas, Isabelle Régner, Jean-Marc Monteil, Herb Marsh, Jerry Suls, Ladd Wheeler, and the current author agreed to continue their ongoing collaboration to clarify the relation between social comparison processes and the BFLPE and to identify possible moderating factors. All parties jointly developed a questionnaire that was administered to French high school students in 2005. This questionnaire formed the basis of the current study.

There were limitations inherent in Study 2 due to its very nature (a further analysis of previously published data that were not designed to examine the BFLPE). These limitations concerned the lack of measures normally used to demonstrate a BFLPE, namely, a standardised academic achievement test that provided a common metric for comparing individual student ability, and a multi-item, psychometrically sound scale to measure academic self-concept. Since the focus of the Blanton et al. (1999) and Huguet et al. (2001) studies was on the influence of comparison processes on performance and not on the BFLPE, neither of these measures was present. For this reason, Study 2 used grades in the absence of a standardised test and a single item self-evaluation measure instead of a multi-item academic self-concept scale. Study 3 overcame these shortcomings by using standardised achievement tests and a multi-item, psychometrically sound measure of academic self-concept. By this means, Study 3 represented one of the first attempts to clarify the relation between the BFLPE and selected social comparisons in a single study.

A further focus of Study 3 was on potential moderators of the BFLPE, not previously examined in Study 1. In addition to individual ability, motivational orientations were investigated to identify whether these constructs alleviated or
aggravated the BFLPE. The impetus for this objective was to identify which students benefit most from selective education, with the intention of providing insights to inform intervention policy.

In summary, as the data that formed Study 2 were originally collected to investigate the effects of social comparisons, that study did not contain instruments that are routinely used to test for the BFLPE. Study 3, therefore, used more appropriate instrumentation to elucidate the relation between the BFLPE and selected social comparisons. Additionally, Study 3 identified further constructs that were examined to ascertain if they aggravated or alleviated the negative effects of the BFLPE.

Aims

The primary purpose of Study 3 was to clarify the relation between social comparison processes and the BFLPE in the context of a single study that addressed previous limitations and to identify possible moderating factors of the BFLPE not examined in Studies 1 or 2. More specifically, Study 2 aimed to:

1. Test the psychometric properties of the instrumentation utilised to measure academic self-concept and achievement goal orientation in the current study in order to demonstrate the robustness of the measures employed;
2. Utilise standardised achievement tests and psychometrically sound academic self-concept scales to test for the presence of the BFLPE;
3. Ascertain whether selected individual social comparisons moderate the negative effects of the BFLPE in order to elucidate the relation between those social comparison processes and the BFLPE;
4. Investigate the nature of the relation between the BFLPE and individual student ability levels in order to explicate whether the BFLPE varies as a function of individual ability; and
5. Determine whether individual differences in achievement goal orientation moderate the BFLPE in order to identify characteristics of individuals that may moderate the BFLPE.
Statement of the Hypotheses and Research Questions

Where previous theory or research provided clear guidance, hypotheses were devised; otherwise research questions were formulated. The numbers associated with the hypotheses and research questions relate to the aim from which they are derived. So, for example, Hypothesis 1.2 is the second hypothesis derived from the first aim, while Research Question 3.1 is the first research question derived from the third aim.

Psychometric Properties of Instrumentation

Hypothesis 1.1: Reliability of the Academic Self-Description Questionnaire II (ASDQ II)

The French and math self-concept scales will be reliable measures of these constructs.

Hypothesis 1.2: Reliability of the Achievement Goal Questionnaire (AGQ)

The AGQ will demonstrate reliability for all four scales measured (mastery approach, mastery avoidance, performance approach, performance avoidance).

Hypothesis 1.3: Factorial Structure of the ASDQ II and the AGQ

The two a priori factor structure of the ASDQ II (French and math self-concept) and the eight a priori factor structure of the AGQ (mastery approach, mastery avoidance, performance approach, and performance avoidance, for French and math) will be verified by confirmatory factor analysis. Figure 6.1 demonstrates the hypothesised two-factor structure of the ASDQ II and the hypothesised eight-factor structure of the AGQ. The latent constructs, or factors, of the ASDQ II are each derived from six observed indicators, resulting in 12 factor loadings. Each of the factors of the AGQ is derived from three indicators, resulting in 24 factor loadings. Straight arrows from each factor to its observed indicators illustrate these relations. In Figure 6.1, factors are depicted as ovals and the items from which they are derived as rectangles. Curved lines indicate correlated uniquenesses between items (see Methodology section). For clarity, correlations between factors were not included in Figure 6.1, but were incorporated in the actual model.
Figure 6.1. Predicted Factor Structure of ASDQ II and AGQ
**Tests of the BFLPE**

*Hypothesis 2.1: Tests of the BFLPE – The Relation Between Individual Ability and Academic Self-Concept in French and Math*

The relation between individual ability, as measured by a subject-specific standardised achievement test, and academic self-concept, as measured by a subject-specific academic self-concept scale (the ASDQ II), will be significantly positive. This prediction will hold for both French and math self-concepts.

*Hypothesis 2.2: Tests of the BFLPE – The Relation Between Class-Average Ability and Academic Self-Concept in French and Math*

Controlling for individual ability in a specific subject, the relation between class-average ability (as measured by the mean standardised test score for each class) and academic self-concept will be significantly negative (the BFLPE; see Figure 2.3). This prediction will hold for both French and math self-concepts.

**Comparison Choice**

*Research Question 3.1: The Relation Between Comparison Choice (Objective Measure) and Academic Self-Concept in French and Math*

What is the relation between comparison choice, measured objectively by the comparison person’s T2 grade, and French and math self-concepts?

*Hypothesis 3.1: Tests of the Moderating Effect of Comparison Choice (Objective Measure) on the BFLPE in French and Math*

The BFLPE will not be moderated by comparison choice measured objectively. For both French and math, the relation between the class-average ability X objective comparison choice interaction and academic self-concept will be small in size and/or not statistically significant.
**Research Question 3.2: The Relation Between Comparison Choice (Subjective Measure) and Academic Self-Concept in French and Math**

What is the relation between comparison choice, as measured by a subjective comparative evaluation measure, and French and math self-concepts?

**Research Question 3.3: Tests of the Moderating Effect of Comparison Choice (Subjective Measure) on the BFLPE in French and Math**

Will the BFLPE be moderated by comparison choice measured subjectively? What is the relation between the subjective comparison choice X class-average ability interaction and French and math self-concepts?

**Individual Ability**

**Research Question 4.1: Tests of the Moderating Effect of Individual Ability on the BFLPE in French and Math**

Is the BFLPE moderated by individual ability level? What is the relation between the individual ability X class-average ability interaction and French and math self-concepts?

**Achievement Goal Orientation**

**Hypothesis 5.1: Relation between Achievement Goal Orientation and Academic Self-Concept in French and Math (Mastery and Performance Approach)**

The relation between mastery and performance approach goal orientations, as measured by the AGQ (Elliot & McGregor, 2001), and academic self-concept for French and math will be significantly positive.
**Hypothesis 5.2: Relation between Achievement Goal Orientation and Academic Self-Concept in French and Math (Mastery and Performance Avoidance)**

Mastery and performance avoidance goal orientations, as measured by the AGQ (Elliot & McGregor, 2001), will have significantly negative associations with French and math self-concepts.

**Research Question 5.1: Tests of the Moderating Effect of Achievement Goal Orientation on the BFLPE in French and Math**

Can individual differences in achievement goal orientation moderate the negative effect of class-average ability on academic self-concept? In particular, what is the relation between the achievement goal orientation X class-average ability interaction and French and math self-concepts?

**Rationale for Hypotheses and Research Questions**

**Psychometric Properties of Instrumentation**

**Rationale for Hypotheses 1.1 and 1.3: Reliability and Factor Structure of the ASDQ II**

The current study utilised the French and math self-concept scales of the ASDQ II (Marsh, 1990b, 1992), which is based on the Marsh/Shavelson model described in Chapter 2. Byrne (1996) reviewed the ASDQ II favourably, stating that it provided “for a critically important means for testing for multidimensional academic self-concepts for adolescents” (p. 130). The ASDQ II has been validated empirically with calls being made for researchers to use self-concept scales specific to subject areas (e.g., Marsh, 1990b; 1992; Marsh & Yeung, 1997a; 1997b). Marsh (1990b) demonstrated that the scales of this instrument have high reliabilities ranging from .885 to .949 (median = .921). Testing of the psychometric properties of the ASDQ II has shown it to be a sound multidimensional measure of academic self-concept, tapping self-concepts for 15 academic subjects (Marsh, 1990b). Factor loadings for each of the items were high and statistically significant, ranging from .668 to .967 (median = .889) and goodness-of-fit indices were also acceptable (TLI = .939). It is anticipated that the current study will display similar results regarding the
psychometric properties of the ASDQ II. Hence, the French and math self-concept scales of the ASDQ II in the current study are expected to demonstrate high reliabilities and the hypothesised a priori factor structure is expected to be confirmed.

**Rationale for Hypotheses 1.2 and 1.3: Reliability and Factor Structure of the AGQ**

The AGQ (Elliot & McGregor, 2001) was developed to assess four factors of achievement motivation: mastery approach goals, mastery avoidance goals, performance approach goals, and performance avoidance goals. Internal reliability for these factors has been shown to be high: .83 for performance avoidance goals; .88 for mastery avoidance goals; .89 for mastery approach goals; and .94 for performance approach goals (Elliot & McGregor, 2001). Research assessing the psychometric properties of this instrument has supported the hypothesised structure (e.g., Conroy & Elliot, 2004; Conroy, Elliot, & Hofer, 2003; Elliot & Reis, 2003; Finney, Pieper, & Barron, 2004; Karabenick, 2003). For example, in testing the 2 X 2 factorial structure of the AGQ, Elliot and McGregor (2001) noted acceptable goodness-of-fit statistics (RMSEA = .042; TLI = .99; and CFI = .99). Additionally, they observed factor loadings ranging from .64 to .98, and intercorrelations ranging from -.09 to .32. Hence, for both math and French in the current investigation, it is expected that the scales of the AGQ will demonstrate reliability and that the hypothesised a priori factor structure will be supported.

**Tests of the BFLPE**

**Rationale for Hypothesis 2.1: Tests of the BFLPE – The Relation Between Individual Ability and Academic Self-Concept in French and Math**

As noted in Chapter 2, research has consistently demonstrated a statistically significant positive relation between individual ability and various facets of academic self-concept (e.g., Marsh, 1994; Marsh & Hau, 2003; Marsh et al., 2000; Marsh & Parker, 1984; Marsh et al., in press). Thus, it seems appropriate to conclude that a similar relation will exist between individual ability and both math and French self-concepts in the current study.
Rationale for Hypothesis 2.2: Tests of the BFLPE – The Relation Between Class-Average Ability and Academic Self-Concept in French and Math

As noted in Chapter 2, many studies have shown that class-average ability has a negative association with academic self-concept (e.g., Craven et al., 2000; Marsh, 2005; Marsh & Craven, 2002; Marsh & Hau, 2003; Marsh et al., 2001; Marsh & Rowe, 1996; Marsh, et al., in press; Mulkey et al., 2005). For this reason it is predicted that a similar pattern of results will be demonstrated in the present investigation.

Comparison Choice

Rationale for Research Question 3.1: The Relation Between Comparison Choice (Objective Measure) and Academic Self-Concept in French and Math

In all but one instance, both Blanton et al. (1999) and Huguet et al. (2001) found that the comparison person’s grade (an objective measure of comparison) was not related to self-evaluation. However, in Study 2 of the present thesis (see Chapter 5), analyses of the Huguet et al. expanded data found a statistically significant positive association between comparison choice (measured objectively) and self-evaluation for two school subjects (French and math). As the evidence is conflicting, and as past research is therefore unable to provide guidance for formulating clear predictions, a research question is posed to illuminate this question.

Rationale for Hypothesis 3.1: Tests of the Moderating Effect of Comparison Choice (Objective Measure) on the BFLPE in French and Math

In Study 2 of this thesis, further analyses of the Blanton et al. (1999) and Huguet et al. (2001) expanded data found that objective comparison choice did not moderate the BFLPE in 9 of 10 tests (see Chapter 5). There was a significant moderating effect of comparison choice on the BFLPE for one school subject only (Dutch). The conclusion drawn from these results was that objective comparison choice did not moderate the BFLPE and it is therefore expected that similar results will be obtained in the current study.
Rationale for Research Question 3.2: The Relation Between Comparison Choice (Subjective Measure) and Academic Self-Concept in French and Math

In addition to the objective measure, the current study also used a subjective evaluation measure to examine comparison choice. Social comparison research has demonstrated that subjectively comparing one’s achievements with a better performing other can either raise self-evaluations, creating an assimilation effect (e.g., Burleson et al., 2005; Collins, 1996, 2000; Lockwood & Kunda, 1997; Wheeler & Suls, 2007), or lower them, creating a contrast effect (e.g., Cash et al., 1983; Morse & Gergen, 1970; Suls & Wheeler, 2000; Thornton & Moore, 1993). For instance, in a related vein, Marsh et al. (in review; see Study 2), found that selecting a more able comparison target had a negative effect on math agency. However, comparing oneself with a less able comparison target has also been shown to produce assimilation (e.g., Brewer & Weber, 1994; Stapel & Suls, 2004) and contrast effects (e.g., Morse & Gergen, 1970; VanderZee et al., 1996). As previous results have been contradictory, a research question is presented to elucidate this issue.

Research Question 3.3: Tests of the Moderating Effect of Comparison Choice (Subjective Measure) on the BFLPE in French and Math

Study 2 of this thesis examined the moderating effect of an objective measure of comparison choice on the BFLPE. However, there is no research of which this author is aware that has examined the moderating effect of a subjective measure of comparison choice on the BFLPE. For this reason, a research question is posed to illuminate this question.

Individual Ability

Rationale for Research Question 4.1: Tests of the Moderating Effect of Individual Ability on the BFLPE in French and Math

As detailed in Chapter 2, evidence surrounding the moderating effect of individual ability on the BFLPE has been conflicting. Additionally, although individual ability moderated the BFLPE in Studies 1 and 2 of the present thesis, the size of the effects was small and hence it was concluded that the BFLPE generalised
well over different ability levels. As the evidence is inconclusive, no a priori predictions are made; instead, a research question is posed to clarify this matter.

**Achievement Goal Orientation**

*Rationale for Hypotheses 5.1 and 5.2: Relation Between Achievement Goal Orientations and Academic Self-Concept in French and Math*

Motivation has been associated with positive educational outcomes (Lepper, 1988) and is considered to be an important predictor of academic performance (Eccles & Wigfield, 2002; Martin, 2001). Early research in achievement goal motivation revealed that, whereas mastery goals were related to positive outcomes, the effects of performance goals were mixed but were more likely to be associated with negative outcomes (Elliot, 2005). However, research using a trichotomous framework of performance approach, performance avoidance, and mastery goals, demonstrated that mastery goals were associated with intrinsic motivation and performance approach goals facilitated performance. Performance avoidance goals were detrimental to both intrinsic motivation and performance (Elliot & Church, 1997). Subsequent tests of the trichotomous model have demonstrated that performance in math and verbal intelligence tests were undermined by performance avoidance goals relative to performance approach and mastery goals (Elliot et al., 2005, Studies 1A and 1B).

Achievement motivation has been also been assessed using a 2 X 2 achievement goal framework consisting of mastery approach, mastery avoidance, performance approach, and performance avoidance goals (Elliot & McGregor, 2001). A test of this framework revealed that whereas performance approach goals were positively associated with academic achievement, performance avoidance goals displayed a negative association. However, no significant relation was found between either mastery approach or mastery avoidance goals with academic achievement (Elliot & McGregor, 2001).

Although no studies have explicitly examined the consequences of holding a specific achievement goal orientation on academic self-concept, findings from previous research would suggest that approach goals will be associated with positive
outcomes and avoidance goals with negative outcomes. Hence, it seems reasonable to predict that mastery and performance approach goals will be positively related to academic self-concept, whereas mastery and performance avoidance goals will have a negative association.

Rationale for Research Question 5.1: Tests of the Moderating Effect of Achievement Goal Orientation on the BFLPE in French and Math

As the moderating effects of achievement goal orientations on the BFLPE have not been previously examined, the nature of their effects can only be speculative. For example, it could be argued that students in high-ability classes with a performance approach goal orientation, because they perform well may be proud of their accomplishments and as a result be untroubled by the successes of others. Consequently, these students may not suffer the negative consequences of the BFLPE. Alternatively, as there is a very high social comparison component in a performance approach goal orientation (see example item in Methodology section), students in high-ability classes perhaps may find the comparisons so overpowering that they suffer the BFLPE more. Mastery goals have been associated with intrinsic motivation, and so it could be speculated that students in high-ability classes who adopt a mastery approach goal orientation may be more focused on their learning and so may not consider the performances of their peers when constructing their self-concept. This focus on learning rather than peers’ performances, while not reducing the BFLPE, certainly should not increase it.

As regards mastery avoidance and performance avoidance goal orientations, students in high-ability classes who hold these goal orientations may be so anxious about their performance that their self-concepts may suffer as a consequence, with a resulting increase in the BFLPE. As the present study is the first to investigate whether one’s achievement goal orientation can moderate the BFLPE, these can be speculations only. Hence, to clarify this matter, a research question is presented.
Methodology

Participants

Participants were 2,015 French students in their first year of high school in 99 classrooms across 16 schools (for an overview of the French educational system, see Chapters 2 and 5). Each school contained between 4 to 10 classes, and there were 9 to 28 students in each class. There were 989 females and 1,023 males. Gender data were missing for three students.

Missing Data

Missing data constitute a pervasive problem for most questionnaire style studies, and the current study was no different. Important aspects of this study included its use of standardised achievement tests, a measure of academic self-concept, and measures of social comparison. However, in all these areas there were missing data. As some students were absent on the day of testing, 90 participants (4.47%) had no standardised test data for math, and 96 (4.76%) were missing standardised French test data. Having students specify with whom they compared their marks allowed objective achievement information for comparison targets to be obtained from standardised tests. Hence, for each student the standardised test score for his/her comparison target was acquired. However, this comparison information was not available for a substantial number of students for three reasons: (1) students could opt not to nominate a comparison target; (2) standardised test data were missing for some students; and (3) the comparison choice could not be identified. This resulted in 333 participants (16.53%) with missing objective comparison data for math and 394 for French (19.55%). Academic self-concept and subjective comparison choice were assessed as part of a questionnaire. There were minimal missing data for the academic self-concept scales ($M = 0.43, SD = 0.26, No. of items = 12$) and for the subjective comparison choice items ($M = 0.47, SD = 0.03, No. of items = 2$), with the average percentage of missing data for all the questionnaire items used in the current study being 0.34 ($SD = 0.19, No. of items = 38$). However, the large amount of missing data for the standardised tests and objective comparison choice data necessitated that this problem be addressed.
Many methods exist to overcome missing data problems. However, methods such as using the last observed value for a participant (longitudinal data), replacing missing values with means calculated from available data, deleting all cases with missing items, and substituting a predicted mean by regressing the incomplete variables on the complete variables, are not only inefficient, but are fraught with problems (Carpenter & Kenward, 2007). Additionally, as is the case with many statistical packages, the statistical package used to analyse these data (MLwiN) uses listwise deletion of non-completers in generating results. This can result in the loss of a large amount of data and consequently produce biased results (du Toit & Mels, 2007). Hence, it was necessary to use a strategy that overcame problems associated with missing data efficiently, without introducing bias. Researchers (e.g., Schafer & Graham, 2002; Tabachnick & Fidell, 2007) suggest that multiple imputation of missing data is superior to more traditional methods.

Multiple imputation uses the Expectation Maximisation algorithm and generates random draws from probability distributions by means of the MCMC algorithm (du Toit & Mels, 2007). Multiple data sets are generated using this procedure. Subsequently, analyses are conducted separately on each data set and results combined for reporting purposes. Moreover, the variance between imputations and within imputations can be calculated providing “a measure of the true uncertainty in the data set caused by the missing data” (Tabachnick & Fidell, 2007, p. 69). Hence, multiple imputation was used to impute data for the missing items in the current study. Consequently, all means and standard deviations reported for the measures used in the current study are based on the imputed data sets. The procedure section below and Appendix A detail how these statistics were computed.

Measures

Individual Ability

Individual ability was measured using standardised tests in math and French. The National Evaluations take place in France at the beginning of each school year (September). These national tests of French and mathematical ability were used to provide a measure of academic achievement that was comparable across all students.
in the study. Scores on these tests ranged from 0 to 100. The average score for the French test was 60.58 ($SD = 18.78$), and for the math test the average score was 59.83 ($SD = 20.92$). These scores were standardised ($M = 0$, $SD = 1$) for analyses.

**Comparison Choice – Objective Measure**

Students nominated the classmate in math and French with whom they preferred to compare their marks. Participants were instructed to leave these items blank if they did not compare with anyone. The comparison person’s scores on the standardised tests in math ($M = 63.05$, $SD = 21.26$) and French ($M = 62.94$, $SD = 19.03$) were used as an objective measure of comparison choice. These scores were subsequently standardised for analyses ($M = 0$, $SD = 1$). The individual ability and objective comparison choice measures were moderately correlated (French: $r = .62$; math: $r = .56$).

**Comparison Choice - Subjective Measure**

Using a 5-point Likert scale, ranging from 1 (much worse) to 5 (much better), with a midpoint of 3 (the same), students were explicitly asked to rate their achievement in French and math compared to the person they nominated as their comparison target. If a student were to respond much worse to this item, then this would be an extremely upward comparison. Alternatively, if a student were to respond much better to this item, then this would be an extremely downward comparison. If students did not nominate anyone with whom they compared themselves, they were instructed to tick a box. Subsequently, these students were given a score of zero (chose not to compare) on this subjective comparison choice measure. Excluding those students who chose not to compare themselves with anyone, the average rating for French was 2.97 ($SD = 1.04$) and for math was 3.06 ($SD = 1.01$). As with the Blanton et al. (1999) and Huguet et al. (2001) data that were the focus of Study 2, students’ scores on this measure tended to cluster around the midpoint of three. In multilevel regression analyses, this variable was treated as categorical with six levels: no choice, much worse, slightly worse, the same, slightly better, and much better.
Academic Self-Concept

The Academic Self-Description Questionnaire II (ASDQ II; Marsh, 1990b, 1992) can be used to measure up to 16 latent constructs. These latent constructs are self-concepts in academic subjects. Only two academic subjects were assessed in this study, French and math self-concept. The wording of items is the same across all the scales, except for the particular subject area. One negatively worded item in each scale was reverse coded for scoring. Full details of this instrument are presented in Table 6.1. The mean scale scores for math and French self-concept suggest that, on average, students rated their academic accomplishments to be above average (see Table 6.2).

Achievement Goals

Motivation was assessed using Elliot and McGregor’s (2001) Achievement Goal Questionnaire (AGQ). This instrument measures four factors: mastery-approach, mastery-avoidance, performance-approach, and performance-avoidance. In the current study, these four latent constructs were assessed for French and math, resulting in eight latent constructs for the AGQ. Mastery approach goals reflect a striving to learn, develop, and improve skills, whereas mastery avoidance goals reflect an anxiety about not learning or understanding all necessary material. Performance approach goals are focused on achievement and performing well compared to others, whereas performance avoidance goals are directed towards the avoidance of failure and incorporate a sense of fear about performing poorly. This instrument was repeated for math and French. Example items, number of items and details of response scales are presented in Table 6.1. Means and standard deviations for each of the four scales for math and French are presented in Table 6.2.

Design

Standardised achievement tests in French and math were administered to students in September 2004 as part of the French National Evaluations. In April 2005, students completed a questionnaire containing, among other items, scales measuring academic self-concept, achievement goal orientation, and subjective comparative
evaluation (see Appendix C). At this time students were also asked to nominate the person with whom they preferred to compare their marks.

Table 6.1. Summary of Key Features of Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>No. of Items</th>
<th>Example Item</th>
<th>Response Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Academic Self-Description Scale II</td>
<td>6</td>
<td>“Compared to others your age you are good at X (math or French).”</td>
<td>6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree)</td>
</tr>
<tr>
<td>(ASDQII)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achievement Goal Questionnaire (AGQ)</td>
<td>12</td>
<td>“In X (math or French) you want to learn as much as possible in class.”</td>
<td>6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree)</td>
</tr>
<tr>
<td>(Mastery Approach)</td>
<td></td>
<td>(Mastery Approach)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“In X (math or French) you worry that you may not learn all that you possibly can in class (Mastery Avoidance)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“In X (math or French) it is important for you to do well compared to others in your class (Performance Approach)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“In X (math or French) your fear of performing poorly in class is often what motivates you (Performance Avoidance)</td>
<td></td>
</tr>
</tbody>
</table>

*Note. a Participants were asked: “Do you agree with the following sentences?” The wording of the original scales was changed to reflect this. b The response scales were altered from the original to allow similar response scales for all measures in this study.*
Table 6.2. Scale Means, Standard Deviations, and Reliability

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean</th>
<th>SD</th>
<th>Cronbach’s α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Self-description Questionnaire II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>3.75</td>
<td>1.08</td>
<td>.89</td>
</tr>
<tr>
<td>Math</td>
<td>4.05</td>
<td>1.07</td>
<td>.88</td>
</tr>
<tr>
<td>Mastery Approach</td>
<td>4.97</td>
<td>0.90</td>
<td>.72</td>
</tr>
<tr>
<td>Mastery Avoidance</td>
<td>4.01</td>
<td>1.15</td>
<td>.68</td>
</tr>
<tr>
<td>Performance Approach</td>
<td>3.88</td>
<td>1.46</td>
<td>.86</td>
</tr>
<tr>
<td>Performance Avoidance</td>
<td>4.44</td>
<td>1.09</td>
<td>.59</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery Approach</td>
<td>5.07</td>
<td>0.82</td>
<td>.64</td>
</tr>
<tr>
<td>Mastery Avoidance</td>
<td>3.99</td>
<td>1.13</td>
<td>.66</td>
</tr>
<tr>
<td>Performance Approach</td>
<td>4.08</td>
<td>1.36</td>
<td>.80</td>
</tr>
<tr>
<td>Performance Avoidance</td>
<td>4.53</td>
<td>1.01</td>
<td>.45</td>
</tr>
</tbody>
</table>

Note. *As noted in Table 6.2, all items were answered on 1-6 Response Scale. SD = Standard deviation.

Procedure

Overview

Multiple Imputation was conducted using the multiple imputation procedure in LISREL (v.8.72). It was considered that five imputed data sets were adequate (Rubin, 1996). Firstly, a Confirmatory Factor Analysis (CFA) was conducted on the imputed data sets simultaneously to verify the structure of the ASDQ II and the AGQ using LISREL. The data sets were then transferred to SPSS (v.14.0) for data manipulation, including scale creation, reliability testing, standardisation, and calculation of class-average ability. These manipulations were implemented on each data set separately. Subsequently, all five data sets were transferred to MLwiN.
where analyses were performed on each data set separately and results combined. To achieve this, using Excel (2000), parameter estimates were averaged, but standard errors were calculated to accommodate variance between and within imputations, using steps outlined by Raudenbush et al. (2005); see Appendix A for details of these steps. These calculations resulted in not only averaged parameter estimates but also standard errors that reflected variance both within and between the imputed data sets. These operations were also applied to the calculation of all reported means and standard deviations. In addition, reliability estimates, factor loadings, and correlations were averaged across the five data sets.

Scale Creation

The factor structures of the ASDQ II and the AGQ were confirmed using CFA via LISREL (v.8.72) across all imputed data sets simultaneously. The statistical package SPSS (v.14.0) was then used to create the ASDQ II and the AGQ scales according to their authors’ instructions for all five data sets. Means and standard deviations for each scale in each data set were calculated and combined as previously described. Reliabilities for the current sample were calculated by averaging results across all five imputed data sets.

Standardisation

For each imputed data set, the National Evaluation test scores were standardised \((M = 0, SD = 1)\) across the entire sample using the statistical package, SPSS (v.14.0), and class-average test scores were calculated by taking the average of students’ standardised scores on this test (French National Evaluations) in every class for French and math. These class-average test scores were not restandardised so that individual student test scores and class-average test scores used a common metric. Test scores for comparison choice, the ASDQ II, and AGQ scales were also standardised across the entire sample. Data were then transferred to the multilevel modelling statistical program, MLwiN (v.2.02), to test hypotheses and explore research questions. In MLwiN, for math and French, cross-products were calculated, but were not restandardised to keep them in the same metric as other measurements.
Statistical Analyses

The current study used a variety of statistical tools to test its hypotheses and research questions. Two of these techniques, confirmatory factor analysis and reliability analysis, are described below. As multilevel modelling was described in detail in Chapter 4, only a summary of its use in the current study is provided.

Confirmatory factor analysis. Very often in psychological research it is not possible to directly measure a construct that one is interested in exploring. Such constructs are often called unobserved or unmeasured variables, latent variables, latent constructs, or factors (Ullman, 2006). For example, teacher morale may not be directly measured, but indicators of morale, such as the team spirit within the school and the amount of pride and enthusiasm felt within the school, can be (see Rowe, 2005). To measure latent constructs that would otherwise be unmeasurable, researchers use previous research and theory to design indicator items, such as those described to test for teacher morale. In the present study, academic self-concept and achievement goal orientation were latent constructs that were measured using published scales that displayed well-defined factor structures in the samples to which they had been applied. However, it was important to determine that these previously demonstrated factor structures were supported in the current sample. For this reason, CFA statistical techniques were employed to verify the factor structure of the scales used in this study.

CFA, performed through structural equation modelling, involves the researcher making a priori predictions, based on substantive theoretical or empirical grounds, about the nature of the relation between indicator variables and a specified latent construct. In the current study, the indicator variables were questionnaire items designed to measure the latent constructs, academic self-concept, and achievement goal orientation. Once the relation between the observed variables and the latent construct is defined, the model is estimated, and examined to ascertain if the proposed model is a good representation of the observed data. In estimating the model, CFA uses maximum likelihood estimation to find parameter values that are as close as possible to the observed data. This is an iterative process that is repeated until the model converges, that is, until it has reached the best possible solution (Brown, 2006).
After specifying and estimating the model, the next step is to determine whether the model is a “good” model and adequately fits the data. Numerous measures of model fit are available. $\chi^2$ is the classic goodness-of-fit index, but it is rarely used in isolation as it is inflated by sample size and it requires that the model holds exactly in the population. A statistically significant $\chi^2$ signifies that the model is not a good fit for the data, but models based on large samples will be routinely rejected on the basis of $\chi^2$. In the current study, three additional goodness-of-fit indices were used to evaluate model fit. The root mean square error of approximation (RMSEA) is less stringent than $\chi^2$, in that the model is not required to hold exactly in the population, but can be a “reasonable” fit. RMSEA values of less than or equal to .05 indicate a close fit, but values between 0.05 and .08 are acceptable. Comparative fit indices (CFI; NNFI) evaluate the fit of the hypothesised model compared to a baseline model, often a null or independence model in which no relations are specified between the variables. Values greater than .95 indicate a model that fits well, but values greater than .9 are considered acceptable (Brown, 2006; Holmes-Smith, 2000).

In addition to goodness-of-fit indices, programs that provide CFA (e.g., LISREL) also produce parameter estimates of factor loadings and factor correlations that are of importance to the current study. The factor loadings demonstrate the relation between the measured indicators and the latent constructs; the factor correlations show relations between the constructs and can be used to demonstrate discriminate validity between scales (Marsh, 1996).

In the current study, each directly measured variable was allowed to load only on its respective factor. Additionally, as the ASDQ II and the AGQ were repeated for French and math this resulted in there being items that had parallel wording, with the only difference being the name of the academic subject. In order to estimate relations between constructs more accurately, and consistent with recommendations by Marsh and Hau (1996), correlated uniquenesses were also incorporated in the model (see also Bong, 2001; Marsh et al., 1999).

**Reliability of scales.** Previous testing of the ASDQ II and the AGQ demonstrated that both display reliability. However, it was important to confirm in the present sample that all items in each scale were measuring a common trait. Hence, reliability
of the ASDQ II and the AGQ was assessed using a measure of internal consistency, which pertains to the extent to which all items in a scale are measuring the same underlying construct. Traditionally, Cronbach’s coefficient alpha has been used as a measure of the internal consistency reliability of a scale (Briggs & Cheek, 1986; Cronbach, 2004) and is calculated by estimating the average correlation among all of the scale items (Pallant, 2001). Values on this statistic range from 0 to 1, with 0 representing no consistency among items and 1 representing complete consistency, and therefore, greater reliability. As a rule of thumb, reliability estimates of .80 and higher are regarded as indicating moderate to high levels of reliability, estimates of .70 are regarded as low, and lower than .60 as unacceptably low (Murphy & Davidshofer, 2001). However, Cronbach’s alpha is directly related to the number of items in the scale. So, scales containing a small number of items will have lower Cronbach’s alpha estimates (Briggs & Cheek, 1986), an especially important consideration for the current study as the AGQ scales contained only three items each.

**Multilevel modelling.** As was the case in the previous two studies, these data had a hierarchical structure: Students were nested within classes and classes were nested within schools. Ignoring this hierarchical structure and using single level statistical techniques can lead to serious problems as outlined in Chapter 4. To overcome these problems, the data in the current study were analysed using a multilevel modelling program (MLwiN, v.2.02) that can deal with data at differing levels. (See Chapter 4 for full details on multilevel modelling.) Hence, multilevel regression analyses were used to analyse these data. Outcome variables, predictor variables, and cross-products used in these analyses are outlined in Table 6.3. In the multilevel analyses conducted, the intercept was allowed to vary across schools, classes and students, but all predictor variables were fixed.

Residuals were inspected at all three levels for every type of analysis conducted in MLwiN. At the school and class levels there was slight evidence of non-linearity and heteroscedasticity for both math and French, but these did not constitute severe violations of assumptions. Accordingly, no action was taken.
Table 6.3. *Summary of Variables Used in Analyses*

<table>
<thead>
<tr>
<th>Outcome Variables</th>
<th>Predictor Variables</th>
<th>Cross-products</th>
</tr>
</thead>
<tbody>
<tr>
<td>French Academic Self-concept; Math Academic Self-concept</td>
<td>Individual ability (Standardised tests in French and math)</td>
<td>Individual ability X class-average ability (French and math)</td>
</tr>
<tr>
<td></td>
<td>Class-average ability (French and math)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comparison target’s ability (Standardised tests in French and math)</td>
<td>Comparison target’s ability X class-average ability (French and math)</td>
</tr>
<tr>
<td></td>
<td>Subjective comparison choice (French and math), a categorical variable with six categories</td>
<td>Subjective comparative evaluation X class-average ability (French and math) (five interactions as one category is the reference group)</td>
</tr>
<tr>
<td></td>
<td>Mastery approach goal orientation (French and math)</td>
<td>Mastery approach goal orientation X class-average ability (French and math)</td>
</tr>
<tr>
<td></td>
<td>Mastery avoidance goal orientation (French and math)</td>
<td>Mastery avoidance goal orientation X class-average ability (French and math)</td>
</tr>
<tr>
<td></td>
<td>Performance approach goal orientation (French and math)</td>
<td>Performance approach goal orientation X class-average ability (French and math)</td>
</tr>
<tr>
<td></td>
<td>Performance avoidance goal orientation (French and math)</td>
<td>Performance avoidance goal orientation X class-average ability (French and math)</td>
</tr>
</tbody>
</table>

*Preliminary Descriptive Statistics*

*Comparison choice – objective measure.* Table 6.4 shows means and standard deviations for individual test scores and comparison choice test scores for French and math. Paired *t* tests, comparing individual test scores and comparison choice test scores (the objective comparison measure), demonstrated that participants chose to compare themselves with a somewhat better performing other in both French and math. This difference was statistically significant for both subjects.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Individual Ability (Test Scores)</th>
<th>Comparison Choice Ability (Test Scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>French</td>
<td>60.58</td>
<td>18.78</td>
</tr>
<tr>
<td>Math</td>
<td>59.83</td>
<td>20.92</td>
</tr>
</tbody>
</table>

Note. SDD = Standard deviation of the difference. df and t values are from paired t tests which were used to determine if there were statistical differences between an individual’s test score and comparison choice’s test score. A tendency toward upward comparison is indicated by positive t values. *p < .001.

**Comparison choice – subjective measure.** Table 6.5 presents means, standard deviations, and paired t tests for individual test scores (standardised) and comparison choice test scores (standardised) for French and math at each level of the subjective comparison measure. Statistically significant upward comparisons in both subjects were made by students who rated themselves as much worse or slightly worse than their comparison targets. Statistically significant downward comparisons in both subjects were made by students who rated themselves as much better than, slightly better than, and similar to their comparison targets. Interestingly, based on the objective comparison choice measure, students tended to compare upward overall (see Table 6.4), but based on the subjective comparison measure students engaged in both upward and downward comparison.
Table 6.5. *Individual Ability, Comparison Choice Ability, and Paired t Tests at Levels of Subjective Comparison Choice (Test Scores Standardised)*

<table>
<thead>
<tr>
<th>Subject</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>SDD</th>
<th>df</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Much Worse</td>
<td>-0.49</td>
<td>0.91</td>
<td>0.15</td>
<td>1.06</td>
<td>1.02</td>
<td>142</td>
<td>7.61**</td>
</tr>
<tr>
<td>Slightly Worse</td>
<td>-0.02</td>
<td>0.98</td>
<td>0.20</td>
<td>0.95</td>
<td>0.84</td>
<td>399</td>
<td>5.46**</td>
</tr>
<tr>
<td>Same</td>
<td>0.12</td>
<td>0.96</td>
<td>0.02</td>
<td>0.96</td>
<td>0.81</td>
<td>684</td>
<td>-3.89**</td>
</tr>
<tr>
<td>Slightly Better</td>
<td>0.20</td>
<td>1.02</td>
<td>-0.17</td>
<td>0.99</td>
<td>0.88</td>
<td>360</td>
<td>-8.46**</td>
</tr>
<tr>
<td>Much Better</td>
<td>-0.05</td>
<td>1.01</td>
<td>-0.64</td>
<td>1.02</td>
<td>0.93</td>
<td>132</td>
<td>-7.59**</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Much Worse</td>
<td>-0.75</td>
<td>0.95</td>
<td>0.18</td>
<td>0.99</td>
<td>0.97</td>
<td>95</td>
<td>10.20**</td>
</tr>
<tr>
<td>Slightly Worse</td>
<td>-0.13</td>
<td>0.99</td>
<td>0.25</td>
<td>0.94</td>
<td>0.86</td>
<td>415</td>
<td>9.27**</td>
</tr>
<tr>
<td>Same</td>
<td>0.07</td>
<td>0.97</td>
<td>-0.02</td>
<td>0.96</td>
<td>0.79</td>
<td>689</td>
<td>-2.98*</td>
</tr>
<tr>
<td>Slightly Better</td>
<td>0.25</td>
<td>0.97</td>
<td>-0.26</td>
<td>0.97</td>
<td>0.82</td>
<td>418</td>
<td>-12.94**</td>
</tr>
<tr>
<td>Much Better</td>
<td>0.11</td>
<td>1.03</td>
<td>-0.68</td>
<td>1.03</td>
<td>0.97</td>
<td>150</td>
<td>-10.42**</td>
</tr>
</tbody>
</table>

*Note.* SDD = Standard deviation of the difference. df and t values are from paired t tests which were used to determine if there were statistical differences between an individual’s test score and comparison choice’s test score. A tendency toward upward comparison is indicated by positive t values. **p < .001; *p < .01.

**Results**

*Overview of Analyses*

Six sets of analyses were conducted to test the hypotheses and research questions posed to satisfy the aims of the current study. The first two sets of analyses consisted of tests of the reliability (Hypotheses 1.1 and 1.2) and factorial structure of the ASDQ II and the AGQ (Hypothesis 1.3). The structure of the ASDQ II and the AGQ was verified using CFA analysis in LISREL. Reliability testing was conducted for all scales in SPSS.

Next, multilevel regression analyses were conducted to test for the presence of the BFLPE (Hypotheses 2.1 and 2.2). For each academic subject, the outcome variable was self-concept for that subject, and predictors were individual ability, as evidenced by standardised test scores, and class-average ability.

The moderating effect of comparison choice on the BFLPE was then examined using separate multilevel regression analyses for objective and subjective comparison choice (Research Questions 3.1, 3.2, 3.3, and Hypothesis 3.1). When objective
comparison choice was examined, the outcome variable was self-concept for each subject and predictors were individual ability, as evidenced by standardised test scores, class-average ability, the standardised test scores of the comparison choice, and the interaction of class-average ability and the standardised test scores of the comparison choice. Subjective comparison choice was treated as a categorical variable. Hence, when subjective comparison choice was examined as a moderator of the BFLPE, the outcome variable was self-concept for each subject and predictors were individual ability, as evidenced by standardised test scores, class-average ability, the categories of subjective comparison choice, and the interaction of class-average ability with each of the subjective comparison choice categories.

The moderating effect of individual ability on the BFLPE was tested in a fifth set of multilevel regression analyses (Research Question 4.1). The outcome variable was self-concept for each subject and predictors were individual ability, as evidenced by standardised test scores, class-average ability, and the interaction of individual ability and class-average ability.

Lastly, multilevel regression analyses were conducted to examine the moderating effect of achievement goal orientation on the BFLPE (Hypotheses 5.1 and 5.2, and Research Question 5.1). The outcome variable was self-concept for each subject and predictors were individual ability, as evidenced by standardised test scores, class-average ability, each of the achievement goal orientations and the interaction of each achievement goal orientation with class-average ability.

Psychometric Properties of Instrumentation

Results for Hypothesis 1.1: Hypothesis 1.1: Reliability of the ASDQ II

This hypothesis predicted that the French and math scales of the ASDQ II would be reliable. Reliability analyses indicated that the coefficient alpha for the French self-concept scale was .89, and for the math self-concept scale was .88 (see Table 6.2). Reliability scores for both the French and math scales of the ASDQ II were high, providing support for the internal consistency of the ASDQ II instrument. As results were consistent with predictions, Hypothesis 1.1 is accepted.
Results for Hypothesis 1.2: Reliability of the AGQ

Hypothesis 1.2 predicted that for math and French all four scales of the AGQ would be reliable. Results of reliability analyses for the four scales of the AGQ are presented in Table 6.2. For math and French, the performance approach scales demonstrated acceptable reliability (.80 and .86 respectively) and the mastery approach and avoidance scales displayed low, but acceptable, reliability (math mastery approach = .64 and math mastery avoidance = .66; French mastery approach = .72 and French mastery avoidance = .68). Reliability estimates for the performance avoidance scales in math and French were in the unacceptably low region (.45 and .59 respectively).

Hypothesis 1.2 is accepted for the performance approach scales and the mastery approach and avoidance scales of the AGQ as these displayed acceptable reliability, thus providing support for the internal consistency of these scales. Hypothesis 1.2 is not accepted for the performance avoidance scales in math and French, as reliability estimates for these scales were unacceptably low. However, Cronbach’s alpha is directly related to the number of items in the scale and so short scales such as these often display low reliability (i.e., with less than less than 10 items; see Briggs & Cheek, 1986). Hence, it was decided to retain this scale in further analyses, but to treat results cautiously.

Results for Hypothesis 1.3: Factorial Structure of the ASDQ II and the AGQ

Hypothesis 1.3 predicted that the two-factor structure of the ASDQ II and the eight-factor structure of the AGQ would be verified by confirmatory factor analysis. The five imputed files were simultaneously subjected to a CFA, and the resulting factor loadings and correlations were averaged to provide single estimates. Table 6.6 provides results for the CFA analysis of the two-factor structure of the ASDQ II and the eight-factor structure of the AGQ. Items were only allowed to load on their respective factors. Correlated uniquenesses were allowed between items that were worded similarly except for the name of the academic subject.

Inferential $\chi^2$ tests indicated that the model should be rejected, but this test is known to be extremely sensitive to large samples, whereby all large models would be rejected using this test (Bentler & Bonett, 1980; Byrne, 1998). However, fit indices
were within acceptable limits (CFI = .96; NNFI = .95; RMSEA = .050), suggesting that the hypothesised model fit the data well. The factor loading for each item was consistently high on the factor it was intended to measure (with the possible exception of performance avoidance for math, which was also unreliable), suggesting that all 10 factors were well defined. Factor loadings were all statistically significant.

An investigation of the factor correlations (see Table 6.6) suggested that math and French self-concepts were distinct concepts, displaying a low correlation of .27. Correlations between the two self-concept factors and the achievement goal factors were low to moderate, ranging from -.48 to .33 (mean r = .048), suggesting that the two self-concept factors were distinct from the achievement goal factors. The negative correlations were logical in that self-concepts were negatively correlated with avoidance goals. The highest correlations were between achievement goals in the two subject areas (mastery approach French and math, r = .72; mastery avoidance French and math, r = .68; performance approach French and math, r = .89; and performance avoidance French and math, r = .83; mean r = .78). Otherwise, correlations between achievement goals were low to moderate (range = .08 to .52). The average correlation between achievement goals was .39, suggesting that the factors were reasonably distinct. Also of note is the variation in the level of domain specificity between the ASDQ II and the AGQ. The ASDQ II is extremely domain specific as seen in the low correlation between French and math self-concepts (.27). The AGQ appears to be less domain specific as correlations between goals for math and French tend to be higher. Mastery goals are less domain specific than the ASDQ II scales (e.g., correlation between mastery approach in math and French = .72) and there appears to be very little domain specificity in performance goals (e.g., correlation between performance approach in math and French = .89).

Results from the CFA analysis clearly identified the two factors of the ASDQ II, and the eight factors of the AGQ that were predicted. Self-concepts for math and French were clearly defined and shown to be distinct from one another, as were the eight achievement goal factors. The goodness-of-fit indicators and factor loadings were within appropriate ranges, indicating that the model was a suitable fit for the data and the factor correlations provided evidence of discriminative validity between the scales. Given these results, Hypothesis 1.3 is accepted.
Table 6.6. Results of Confirmatory Factor Analyses for ASDQII and AGQ

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.82</td>
<td>.82</td>
<td>.51</td>
<td>.56</td>
<td>.73</td>
<td>.46</td>
<td>.64</td>
<td>.61</td>
<td>.84</td>
<td>.68</td>
</tr>
<tr>
<td>2</td>
<td>.79</td>
<td>.84</td>
<td>.63</td>
<td>.60</td>
<td>.80</td>
<td>.51</td>
<td>.67</td>
<td>.60</td>
<td>.84</td>
<td>.51</td>
</tr>
<tr>
<td>3</td>
<td>.73</td>
<td>.75</td>
<td>.70</td>
<td>.74</td>
<td>.74</td>
<td>.47</td>
<td>.73</td>
<td>.73</td>
<td>.79</td>
<td>.56</td>
</tr>
<tr>
<td>4</td>
<td>.61</td>
<td>.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.84</td>
<td>.86</td>
<td>.73</td>
<td>.74</td>
<td>.79</td>
<td>.56</td>
<td>.73</td>
<td>.73</td>
<td>.79</td>
<td>.56</td>
</tr>
</tbody>
</table>

Factor Correlations

<table>
<thead>
<tr>
<th></th>
<th>MSC</th>
<th>.100</th>
<th>FSC</th>
<th>.27</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>MastApM</td>
<td>.29</td>
<td>.17</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MastAvM</td>
<td>-.48</td>
<td>-.05</td>
<td>.19</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>PerfApM</td>
<td>.19</td>
<td>.08</td>
<td>.37</td>
<td>.08</td>
<td>1.00</td>
</tr>
<tr>
<td>PerfAvM</td>
<td>-.09</td>
<td>.07</td>
<td>.41</td>
<td>.52</td>
<td>.46</td>
</tr>
<tr>
<td>MastApFr</td>
<td>.07</td>
<td>.32</td>
<td>.72</td>
<td>.29</td>
<td>.25</td>
</tr>
<tr>
<td>MastAvFr</td>
<td>-.12</td>
<td>-.33</td>
<td>.29</td>
<td>.68</td>
<td>.14</td>
</tr>
<tr>
<td>PerfApFr</td>
<td>.06</td>
<td>.20</td>
<td>.30</td>
<td>.15</td>
<td>.89</td>
</tr>
<tr>
<td>PerfAvFr</td>
<td>.00</td>
<td>-.02</td>
<td>.27</td>
<td>.31</td>
<td>.43</td>
</tr>
</tbody>
</table>

Note. All factor loadings are presented in completely standardised format. MSC = Math Self-concept; FSC = French Self-concept; MastAppM = Math Mastery Approach; MastAvM = Math Mastery Avoidance; PerfAppM = Math Performance Approach; PerfAvM = Math Performance Avoidance; MastAppFr = French Mastery Approach; MastAvFr = French Mastery Avoidance; PerfAppFr = French Performance Approach; PerfAvFr = French Performance Avoidance; Sample size = 2015; Chi square = 13647.23, degrees of freedom = 531 p = 0.00; Comparative Fit Index = .96; Non-Normed Fit Index = .95; Root Mean Square Error of Approximation = .050.
Summary for Hypotheses 1.1 and 1.2 Reliability and Factor Structure of ASDQ II and AGQ

Results of analyses demonstrated the internal reliability of the ASDQ II and of the performance approach scales of the AGQ. The reliability of the remaining AGQ scales was more problematic. The predicted two-factor structure of the ASDQ II and the eight-factor structure of the AGQ were verified by confirmatory factor analysis.

Tests of the BFLPE

Results for Hypothesis 2.1: Tests of the BFLPE – The Relation Between Individual Ability and Academic Self-Concept in French and Math

Hypothesis 2.1 predicted that there would be a statistically significant positive relation between both French ability and French self-concept and between math ability and math self-concept. Results for Hypothesis 2.1, presented in Table 6.7, indicated that there was a significant positive relation between individual ability and self-concept for French and math (French = .505 and math = .690). Students with higher ability had higher self-concepts. Given that there was a statistically significant positive relation between ability and self-concept, Hypothesis 2.1 is accepted.

Results for Hypothesis 2.2: Tests of the BFLPE – The Relation Between Class-Average Ability and Academic Self-Concept in French and Math

Hypothesis 2.2 predicted that there would be a statistically significant negative relation between class-average ability and self-concept for both French and math. As seen in Table 6.7, when individual ability was controlled, there was a significantly negative relation between class-average ability and self-concept for French (-.515) and math (-.545). Students in classes of higher ability had lower self-concepts. As there was a significant negative relation between class-average ability and self-concept, Hypothesis 2.2 is accepted.

Summary for Hypotheses 2.1 and 2.2: Tests of the BFLPE

Results for Hypotheses 1.1 and 1.2 demonstrated a significantly positive relation between individual ability and self-concept and a significantly negative relation between class-average ability and self-concept for both academic subjects. These results signify that a BFLPE was present for French and math in these data.
As there was a considerable amount of data missing for, among other measures, the comparison target’s test score, multiple imputation was employed to compensate for the missing information. However, knowing that there were participants for whom no comparison information was available, the question arose: Did these participants also display a BFLPE? Table 6.8 offers evidence that a BFLPE existed for those for whom no comparison information was available, although it was somewhat reduced. There was a significantly positive relation between individual ability and class-average ability (French = .394; math = .649) and class-average ability had a significantly negative relation with self-concept (French = -.447; math = -.484).

Table 6.7. The BFLPE for French and Math Self-Concept

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Ability</td>
<td>Class-Average</td>
</tr>
<tr>
<td>French</td>
<td>0.505* (0.026)</td>
<td>-0.515* (0.063)</td>
</tr>
<tr>
<td>Math</td>
<td>0.690* (0.024)</td>
<td>-0.545* (0.059)</td>
</tr>
</tbody>
</table>

*Note. Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). All parameter estimates are standardised coefficients. SE in parentheses.

Table 6.8. BFLPE for Non-Choosers for French and Math Self-Concept

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Ability</td>
<td>Class-Average</td>
</tr>
<tr>
<td>French</td>
<td>0.394* (0.062)</td>
<td>-0.447* (0.093)</td>
</tr>
<tr>
<td>Math</td>
<td>0.649* (0.060)</td>
<td>-0.484* (0.100)</td>
</tr>
</tbody>
</table>

*Note. Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). All parameter estimates are standardised coefficients. SE in parentheses.
Comparison Choice

*Results for Research Question 3.1: Relation Between Comparison Choice (Objective Measure) and Academic Self-Concept in French and Math*

Research Question 3.1 enquired about the relation between objective comparison choice and French and math self-concepts. As seen in Table 6.9, the objective comparison choice measure was statistically significantly positively related to academic self-concept for French (.086) and math (.104), indicating that comparisons with someone performing better were positively associated with students’ academic self-concepts.

*Results for Hypothesis 3.1: Tests of the Moderating Effect of Comparison Choice (Objective Measure) on the BFLPE in French and Math*

Hypothesis 3.1 predicted that the interaction of class-average ability and the objective measure of comparison choice on math and French self-concepts would be small in size and/or not statistically significant. Results indicated that a BFLPE was evident in these data: Individual ability was positively related to self-concept (French = .481; math = .660) and class-average ability was negatively related (French = -.519; math = -.586; see Table 6.9). Results for the objective comparison choice X class-average ability interaction demonstrated a statistically significant relation with French self-concept (.086). This interaction is depicted in Figure 6.2. For math, the interaction effect was not statistically significant (.045). Hence, analyses indicated that objective comparison choice moderated the BFLPE for French, but not for math. Whereas the latter finding for math was as expected, the former result for French was contrary to expectations. Consequently, Hypothesis 3.1 is accepted for math but not for French.

For math, the comparison made no difference to the BFLPE: Students in high-ability classes had lower academic self-concepts irrespective of their comparison target. However, for French the comparison did make a difference. Figure 6.2 depicts this interaction in a graph. Smaller BFLPEs were associated with students who compared with a high-performing other. Students in low-ability classes tended to have similar French self-concepts, irrespective of the achievements of their comparison targets. In general, students in high-ability classes tended to have lower French self-concepts than their counterparts in low-ability classes. However, students
in high-ability classes who compared themselves with a high-performing other had higher academic self-concepts than those who compared themselves with someone who performed less well. Thus, the BFLPE was more pronounced for students who compared themselves with low-performing others.

**Summary for Research Question 3.1 and Hypothesis 3.1: Comparison Choice (Objective Measure)**

The objective comparison measure displayed a positive relation with both French and math self-concepts, such that students who compared themselves with a better performing other had higher self-concepts. The objective comparison measure had a moderating effect on the BFLPE for French only. In this instance, smaller BFLPEs were associated with students who compared themselves with a high-performing other.

<table>
<thead>
<tr>
<th>Table 6.9. Objective Comparison Choice as a Moderator of the BFLPE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Effects</strong></td>
</tr>
<tr>
<td>Subject</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>French</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Math</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Note.* Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). All parameter estimates are standardised coefficients. SE in parentheses. Comp = Comparison Other Ability; Comp X Cls = Comparison other Ability X Class-average Ability; Cons = Constant; Ind = Individual.
**Figure 6.2.** The Effect of the Objective Comparison Choice X Class-Average Ability Interaction on French Self-Concept

*Note.* Figure based on predicted values. High class-average ability = 1 standard deviation above the mean for class-average ability, and low class-average ability = 1 standard deviation below the mean. Similarly, high comparison choice test score = 1 standard deviation above the mean for comparison choice test score, and low comparison choice test score = 1 standard deviation below the mean. Individual ability held constant at mean.

**Results for Research Question 3.2: The Relation Between Comparison Choice (Subjective Measure) and Academic Self-Concept in French and Math**

This research question enquired about the relation between subjective comparison choice and French and math self-concepts. Results are presented in Table 6.10. Compared to students who rated their performance in French as the same as their comparison target’s performance, students who did not choose a comparison target (-.273) and those who thought their performance was much worse (-.849) or slightly worse (-.341) than their target’s performance had significantly lower academic self-concepts. Conversely, students who rated their performance in French as slightly better (.188) or much better (.388) had significantly higher academic self-concepts than those who rated their performance as similar to their comparison target’s performance. Results were similar for math. Compared to those who rated their performance as similar to their target, those who did not choose, (-.241), or who thought they were much worse (-.712) or slightly worse (-.227) than their target had
significantly lower academic self-concepts, whereas those who rated themselves as slightly better (.185) or much better (.503) had significantly higher academic self-concepts.

As regards this subjective measure, recall that students who responded with *slightly worse* or *much worse* to this item would have made upward comparisons. Students who responded with *slightly better* or *much better* to this item would have made downward comparisons. The subjective comparison choice results suggest that, compared to those who made lateral comparisons, those who did not compare and those who compared upward with someone considered to be performing better had lower academic self-concepts. Alternatively, downward comparisons with someone performing less well had a positive effect on one’s academic self-concept compared to lateral comparisons. Importantly, students appear to have been correct in their perceptions of their abilities in relation to their comparison targets. Results of paired *t* tests displayed in Table 6.5 demonstrate that students who rated themselves as much worse or slightly worse than their comparison targets did in fact make upward comparisons and students who rated themselves as much better than, slightly better than, and similar to their comparison targets appear to have made downward comparisons.

**Results for Research Question 3.3: Tests of the Moderating Effect of Comparison Choice (Subjective Measure) on the BFLPE in French and Math**

This research question asked whether the BFLPE would be moderated by comparison choice measured subjectively. In particular, it asked about the relation between French and math self-concepts and the subjective comparison choice X class-average ability interaction. As seen in Table 6.10, a BFLPE was evident: Individual ability was positively related to self-concept (French = .417; math = .607) and class-average ability was negatively related (French = -.413; math = -.504). However, none of the subjective comparative evaluation X class-average ability interactions was statistically significant, for either of the academic subjects tested. Hence, the BFLPE was not moderated by subjective comparison choice. Irrespective of how students rated themselves vis-à-vis their comparison target, students in high-ability classes had lower academic self-concepts than their counterparts in low-ability classes.
Table 6.10. *Subjective Comparison Choice as a Moderator of the BFLPE (“The Same” is the reference category)*

<table>
<thead>
<tr>
<th></th>
<th>French</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual Ability</td>
<td>0.417*</td>
<td>0.607*</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Class-Average Ability</td>
<td>-0.413*</td>
<td>-0.504*</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.071)</td>
</tr>
<tr>
<td>No Choice</td>
<td>-0.273*</td>
<td>-0.241*</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Much Worse</td>
<td>-0.849*</td>
<td>-0.712*</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.090)</td>
</tr>
<tr>
<td>Slightly Worse</td>
<td>-0.341*</td>
<td>-0.227*</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Slightly Better</td>
<td>0.188*</td>
<td>0.185*</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>Much Better</td>
<td>0.388*</td>
<td>0.503*</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>NC X Class</td>
<td>-0.099</td>
<td>-0.100</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.094)</td>
</tr>
<tr>
<td>MW X Class</td>
<td>0.020</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.122)</td>
</tr>
<tr>
<td>SW X Class</td>
<td>0.054</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>SB X Class</td>
<td>-0.008</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>MB X Class</td>
<td>0.012</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.100)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.105*</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.051)</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>0.012</td>
<td>0.022*</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Class</td>
<td>0.064*</td>
<td>0.028*</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Individual</td>
<td>0.696*</td>
<td>0.609*</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>

*Note.* Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). All parameter estimates are standardised coefficients. SE in parentheses. NC X Class = No choice X Class-average Ability; MW X Class = Much worse X Class-average Ability; SW X Class = Slightly worse X Class-average Ability; SB X Class = Slightly better X Class-average Ability; MB X Class = Much better X Class-average Ability.
Summary for Research Questions 3.2 and 3.3: Comparison Choice (Subjective Measure)

Results for the main effect of the subjective comparison measure suggested that students who compared upward had lower academic self-concepts and that students who made downward comparisons had higher academic self-concepts. However, the BFLPE was not moderated by the subjective comparison measure for either French or math.

Appropriateness of Multiple Imputation

As noted previously, there was a large amount of missing data especially on the objective comparison choice dimension (comparison target’s score on standardised test), and so these results were based on multiply imputed data. An important consideration in determining the appropriateness of imputing these data was whether similar results would be achieved were the comparison target’s scores or the subjective comparison choice scores not imputed. Analyses, performed on data where the comparison choice information was not imputed, revealed a similar pattern of results to those reported above and are presented in Appendix C. Therefore, imputing the missing data was an appropriate strategy to adopt.

Individual Ability

Results for Research Question 4.1: Tests of the Moderating Effect of Individual Ability on the BFLPE in French and Math

This research question asked whether the BFLPE could be moderated by individual ability level. Specifically, it examined the relation between the individual ability X class-average ability interaction and math and French self-concepts. As seen in Table 6.11, a BFLPE was evident. Individual ability was positively related to self-concept (French = .522; math = .696) and class-average ability was negatively related (French = -.400; math = -.478). The individual ability X class-average ability interaction had a statistically significant positive association with academic self-concept for French (.200) and math (.141). The pattern of this interaction is portrayed in Figures 6.3 and 6.4.
An inspection of the graph of the interaction suggests that for French the BFLPE was less for students of higher ability (see Figure 6.3). In low-ability classes, students of higher ability had higher French self-concepts than low or average ability students. In high-ability classes, compared to their counterparts in low-ability classes, students of high-ability had slightly lower French self-concepts. However, compared to students of equal ability in low-ability classes, low and average ability students who attended high-ability classes had lower French self-concepts. This was especially the case for low-ability students who suffered the BFLPE to a greater extent than average and high-ability students.

Similarly, for math the BFLPE was more pronounced for students of low-ability (see Figure 6.4). Students of all ability levels in high-ability classes had lower math self-concepts than their counterparts in low-ability classes. However, this reduction in math self-concept was greater for low-ability students. As was the case for French, it was the low math ability students who suffered the BFLPE to a greater extent.

Table 6.11. Individual Ability as a Moderator of the BFLPE

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual Ability</td>
<td>Class - Average X Class</td>
</tr>
<tr>
<td>French</td>
<td>0.522* (0.026)</td>
<td>-0.400* (0.066)</td>
</tr>
<tr>
<td>Math</td>
<td>0.696* (0.024)</td>
<td>-0.478* (0.061)</td>
</tr>
</tbody>
</table>

Note. Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). All parameter estimates are standardised coefficients. SE in parentheses. Ind Ab X Class = Individual Ability X Class-average Ability; Con = Constant; Ind = Individual.
Figure 6.3. The Effect of the Individual Ability X Class-Average Ability Interaction on French Self-Concept

Note. Figure based on predicted values. High class-average ability = 1 standard deviation above the mean for class-average ability, and low class-average ability = 1 standard deviation below the mean. Similarly, high individual ability = 1 standard deviation above the mean for individual ability, and low individual ability = 1 standard deviation below the mean.

Figure 6.4. The Effect of the Individual Ability X Class-Average Ability Interaction on Math Self-Concept

Note. Figure based on predicted values. High class-average ability = 1 standard deviation above the mean for school-average ability, and low class-average ability = 1 standard deviation below the mean. Similarly, high individual ability = 1 standard deviation above the mean for individual ability and low individual ability = 1 standard deviation below the mean.
Results for Hypothesis 5.1: Relation between Achievement Goal Orientation and Academic Self-Concept in French and Math (Mastery and Performance Approach)

Hypothesis 5.1 predicted that mastery and performance approach goal orientations would have a statistically significantly positive relation with academic self-concept for math and French. Results of analyses (see Table 6.12) indicated that mastery approach and performance approach goal orientations had a statistically significant positive association with French self-concept (0.273 and 0.149 respectively) and math self-concept (0.214 and 0.139 respectively). Students who held mastery and performance approach goal orientations had higher self-concepts in both French and math. As these results were as predicted, Hypothesis 5.1 is accepted.

Results for Hypothesis 5.2: Relation between Achievement Goal Orientation and Academic Self-Concept in French and Math (Mastery and Performance Avoidance)

This hypothesis predicted that mastery and performance avoidance goal orientations would have a significantly negative relation with math and French self-concepts. As displayed in Table 6.12, a mastery avoidance goal orientation had a statistically significant negative association with French self-concept (-0.329) and math self-concept (-0.315). A performance avoidance goal orientation was not statistically significantly associated with either academic subject tested.

Students who espoused mastery avoidance goals had significantly lower academic self-concepts, consistent with predictions. However, contrary to expectations, holding a performance avoidance goal orientation was not related to academic self-concept. Given these results, Hypothesis 5.2 is accepted for mastery avoidance, but not for performance avoidance.
Results for Research Question 5.1: Tests of the Moderating Effect of Achievement Goal Orientation on the BFLPE in French and Math

Research Question 5.1 addressed the moderating effect of achievement goals on the BFLPE. It posed the question as to what would be the relation between the achievement goal orientation X class-average ability interaction and academic self-concept for French and math. As seen in Table 6.12, a BFLPE was evident: Individual ability was positively related to self-concept (French = .436; math = .559) and class-average ability was negatively related to self-concept (French = -.401; math = -.386). The interaction effects of class-average ability with mastery approach, mastery avoidance, and performance approach were not statistically significant (see Table 6.12). Adopting mastery or performance approach goals did not moderate the BFLPE. For both French and math, students in high-ability classes who espoused these goal orientations had lower academic self-concepts than their counterparts in low-ability schools. The performance avoidance X class-average ability interaction was statistically significantly negative for French (-0.094), but not for math. Whereas holding a performance avoidance goal in math did not moderate the BFLPE, if students adopted this goal orientation in French the BFLPE was moderated. This interaction is portrayed in Figure 6.5, indicating that the BFLPE was reduced for students who were low in performance avoidance goal orientation. In low-ability classes, students who held a higher performance avoidance orientation had slightly higher French self-concepts than those who were average in performance avoidance. In turn, those who held an average performance avoidance orientation had slightly higher French self-concepts than those who were low in performance avoidance. This relation was reversed in high-ability classes, where those who held a higher performance avoidance orientation had lower French self-concepts than those who were average and low in performance avoidance. Hence, the BFLPE was larger for students who espoused a high performance avoidance goal orientation and lower for students who were low in performance avoidance goal orientation. However, given that this scale was unreliable, results should be treated with caution.
Table 6.12. Achievement Goals as Moderators of the BFLPE

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>French</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Ability</td>
<td>0.436*</td>
<td>0.559*</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Class-Average Ability</td>
<td>-0.401*</td>
<td>-0.386*</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Mastery Approach</td>
<td>0.273*</td>
<td>0.214*</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Mastery Avoidance</td>
<td>-0.329*</td>
<td>-0.315*</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Performance Approach</td>
<td>0.149*</td>
<td>0.139*</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Performance Avoidance</td>
<td>-0.020</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Mastery Approach X Class-Average</td>
<td>-0.011</td>
<td>-0.032</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Mastery Avoidance X Class-Average</td>
<td>-0.039</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Performance Approach X Class-Average</td>
<td>0.001</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Performance Avoidance X Class-Average</td>
<td>-0.094*</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.010</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.036)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>0.000</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Class</td>
<td>0.042*</td>
<td>0.016*</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Individual</td>
<td>0.619*</td>
<td>0.533*</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.017)</td>
</tr>
</tbody>
</table>

Note. Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). All parameter estimates are standardised coefficients. SE in parentheses.
Summary for Hypotheses 5.1, 5.2, and Research Question 5.1: Achievement Goal Orientation

Whereas holding a mastery or a performance approach goal orientation was associated with higher self-concepts in both French and math, holding a mastery avoidance goal orientation was associated with lower self-concepts. Performance avoidance goal orientations were not related to French or math self-concepts. The BFLPE was only moderated by one achievement goal orientation: performance avoidance. Smaller BFLPEs were associated with students low in this goal orientation in French.
Discussion

Overview

The principal aim of the current study was to contribute to clarifying the relation between social comparison processes and the BFLPE, a task initiated in Study 2 of the present thesis. However, due to the very nature of that study – a further analysis of previously published data – the instrumentation used was not appropriate to test for the BFLPE. The present study was undertaken to investigate the nature of the relation between selected social comparisons and the BFLPE using standardised achievement tests and a psychometrically sound measure of academic self-concept, both of which are essential ingredients to appropriately test for the presence of the BFLPE. Consistent with previous research (e.g., Craven et al., 2000; Davis, 1966; Marsh et al., 1995; Marsh & Hau, 2003; Marsh et al., 2001; Marsh & Parker, 1984), a BFLPE emerged in these data. For both academic subjects tested, students in high-ability classes had lower academic self-concepts than their counterparts in low-ability classes. Additionally, analyses indicated that whereas the BFLPE was not moderated by subjective comparison choice, it was moderated by objective comparison choice, but in French only.

The current study examined additional possible moderating factors of the BFLPE, namely, individual ability and achievement goal orientation. Results of analyses demonstrated that larger BFLPEs were associated with students of low-ability and with those students who espoused a high performance avoidance achievement goal orientation.

Psychometric Properties of Instrumentation

The current study began by testing the psychometric properties of the instrumentation utilised to measure academic self-concept and achievement goal orientation, in order to demonstrate the robustness of the measures employed. Consistent with previous research (e.g., Marsh, 1990b), the ASDQ II displayed high reliability, thus offering support for the internal consistency of the instrument. Results for the AGQ were not so clear-cut. Whereas reliability scores for the performance approach scales were high, those of the mastery approach and
avoidance scales were low, but acceptable, and those of the performance avoidance scales were unacceptably low. Hence, these findings differ somewhat from those of previous research that has shown the internal reliability for these factors to be high (Elliot & McGregor, 2001). However, these scales contained only three items each. As Cronbach’s alpha is directly related to the number of items in the scale, low reliability is common with short scales such as these (i.e., with fewer than 10 items; see Briggs & Cheek, 1986). Nevertheless, given the unacceptable levels of reliability for these scales, results pertaining to these scales need to be considered cautiously.

Results from the CFA analysis clearly identified the two factors of the ASDQ II and the eight factors of the AGQ that were predicted. Self-concepts for math and French and the eight achievement goal factors were clearly defined and distinct from each other. Goodness-of-fit indicators and factor loadings indicated that the model was a suitable fit for the data and the factor correlations provided evidence of discriminative validity between the scales. Hence, these results are consistent with previous research (e.g., Elliot & McGregor, 2001; Marsh, 1990b), demonstrating the factor structure of these instruments.

**Comparison Choice as a Moderator of the BFLPE**

Investigating the effect of comparison choice on the BFLPE was a central component of this study and having two measures of comparison choice allowed its effects to be examined from both an objective and a subjective perspective.

**Objective Comparison Choice**

The objective measure was the comparison person’s score on a standardised test, a device that allowed comparison direction to be ascertained without subjectivity. Using this objective measure, preliminary analyses revealed that, on average, students tended to make upward comparisons: They tended to compare themselves with someone who was performing better than they were in both French and math. Tests of the moderating influence of objective comparison choice on the BFLPE indicated that the BFLPE was moderated by choice of comparison target depending on the academic subject assessed: Moderating effects were found for French, but not for math.
For math, objective comparison choice did not moderate the BFLPE. Comparing with someone performing better in math resulted in higher math self-concepts, but this comparison did nothing to reduce the negative effect on math self-concept of attending a high-ability class. Irrespective of their choice of comparison target in math, students in high-ability classes had lower math self-concepts than their counterparts in low-ability classes.

As regards French, smaller BFLPEs were associated with students in high-ability classes who compared with high-performing others. Although attending a high-ability class had a negative effect on French self-concept, if students compared with a high-achieving other they had higher French self-concepts than those who compared with a low-performing other. Students may have felt inspired by their high-achieving comparison targets and have considered that they were “almost as good as the very good ones” (Wheeler, 1966, p. 30), with the result that they were not as affected by BFLPE as students who compared with low-performing targets. These findings are consistent with the social comparison literature, demonstrating that the effects of upward comparisons can be positive and that individuals can assimilate with their better performing targets (e.g., Burleson et al., 2005; Collins, 1996, 2000; Lockwood & Kunda, 1997; Wheeler & Suls, 2007). This result is also consistent with one of the findings of Study 2. When the Blanton et al. (1999) data were re-analysed, results indicated that the BFLPE was moderated, but only for Dutch. Therefore, for both the Blanton et al. re-analysis and the current study, comparison choice moderated the BFLPE, but only for the language of instruction in the particular country in which the study was conducted.

One reason for the difference between the BFLPE moderating results for math and French could be that subjects such as math and science are more discrete subjects than French and social science subjects. (As the present investigation was conducted in France, the study of French would have been similar to that of English in English speaking nations, incorporating elements of both literature and language, giving it the characteristics of a social science subject). Students may be more aware of their standing in math and science, because answers are either correct or incorrect, whereas social science subjects, such as French, are more open to subjectivity in marking. Perhaps when answers to questions set for homework tasks or tests are

308
more clearly defined, then more accurate selected comparisons can be made (see Ireson & Hallam, 2001). Having an accurate appraisal of one’s abilities compared to one’s selected comparison target may mean that the negative effects of attending a high-ability school (the BFLPE) cannot be ignored. Alternatively, one’s standing vis-à-vis one’s selected comparison target may be more vague with social science subjects, where the marking criteria are less rigid, resulting in one feeling that one can aspire to achieve as well as one’s high-achieving comparison target. If this were the case, then results similar to those for math should be seen for science subjects where answers are also clearly defined. Additionally, results similar to those for French should be seen for social science subjects where answers are less clearly defined. Results of the re-analyses of the Blanton et al. (1999) data do not fully support these propositions. In that study, similar to the results for math in the current study, comparing with others did not have a moderating effect on the BFLPE for biology. However, unlike the results for French in the current study, social science subjects (e.g., history, geography) examined in Study 2 also did not display a moderating effect. Given these contradictory findings, the reason for the difference between the BFLPE moderating results for math and French in the current study is still open to conjecture. Additionally, as the moderating effect for French was not replicated in the Huguet et al. (2001) expanded data (the language of instruction for that particular study), these results must await further clarification.

In previous studies (Huguet et al., 2001; see Study 2b of the present investigation) objective comparison choice was positively related to self-evaluation in only a few instances. The original Huguet et al. study found such a positive effect for only one academic subject and Study 2 of this thesis found that comparison choice and self-evaluation were positively related for French and math, but only in the expanded Huguet et al. data. This latter finding was replicated in the current study. In both academic subjects assessed, students selected comparison targets who were performing better than themselves, with corresponding beneficial effects on their academic self-concepts: Students felt more positive about their own academic accomplishments when they compared themselves with someone whose achievement was better than their own. As such, these results are in accordance with previous social comparison findings that upward comparisons can be inspiring and result in
assimilation towards the comparison target (e.g., Burleson et al., 2005; Collins, 1996, 2000; Lockwood & Kunda, 1997; Wheeler & Suls, 2007).

**Subjective Comparison Choice**

Results for subjective comparison choice differed from those of the objective comparison choice measure in two ways. Firstly, whereas objective comparison choice moderated the BFLPE for French, subjective comparison choice did not. For both French and math, students in high-ability classes suffered the negative effects of the BFLPE, irrespective of how they subjectively evaluated their standing vis-à-vis their comparison target. In this respect, these results are consistent with those of Study 2. In that study, the BFLPE was moderated in only one academic subject and so the conclusion was that the BFLPE was not moderated by comparison choice.

Secondly, when measured objectively, upward comparisons had a positive effect on self-concept: Students who chose a better performing comparison target had higher academic self-concepts. But, when measured subjectively, in comparison with lateral comparisons, upward comparisons decreased academic self-concept while downward comparisons increased it. Whereas students who considered that their comparison targets were performing better than they were had lower academic self-concepts, those who considered that their comparison targets were performing at a lower level had higher academic self-concepts. Furthermore, students tended to be correct in their assessment of their achievements vis-à-vis their comparison targets. Consideration of individual students’ test scores and their comparison targets’ test scores indicated that when students rated themselves as worse than their targets, they were objectively worse, and when they rated themselves as better, they were objectively better. Additionally, not comparing also had a negative effect on academic self-concepts. Compared to those who made lateral comparisons, students who reported that they did not compare themselves with others had lower academic self-concepts. Perhaps these students chose not to compare themselves with others because their self-concepts were already low and they did not wish to further damage them by actively making the comparison.

These findings for subjective comparison choice are consistent with social comparison literature that has demonstrated that upward comparisons can be deflating and result in contrast effects (e.g., Cash et al., 1983; Morse & Gergen,
1970; Suls & Wheeler, 2000; Thornton & Moore, 1993) and that downward comparisons can make one feel better about oneself (e.g., Morse & Gergen, 1970; VanderZee et al., 1996; Wheeler & Miyake, 1992; Wills, 1981). They are also consistent with recent BFLPE research that has demonstrated that subjective upward comparisons have a negative effect on math agency (Marsh et al., in review). Furthermore, these findings can be encapsulated within the SA framework. When making a subjective judgment of their achievements relative to their comparison targets, students may have looked for, and found, differences between themselves and their comparison targets. Perhaps finding these differences may have led students to feel good about their accomplishments when making downward comparisons and deflated when making upward comparisons.

**Objective Versus Subjective Comparison Choice**

In the current study upward comparisons (with ostensibly the same person) had differential effects on self-concept depending on whether they were measured objectively or subjectively. Whereas upward comparisons measured objectively led to assimilation, upward comparisons measured subjectively led to contrast. In previous research, subjective and objective response scales have been shown to produce differing results. In particular, subjective scales appear to produce contrast effects and objective scales yield assimilation effects (Biernat et al., 1997; Manis et al., 1991; Mussweiler & Strack, 2000b). Although the objective measure used herein was not a response scale, the differential between the two measures in the current study is reminiscent of this distinction. So, why did the same comparison measured differently have different effects? It could have been that when making their subjective comparisons students simply got it wrong. Maybe upward comparisons were not really upward comparisons at all? However, this was not the case. As already noted, an inspection of the test scores revealed that students were correct in their judgements. Perhaps the differentiation could be due to the subjective variable being categorical and the objective variable being continuous? To test this theory, unreported analyses were conducted using the objective measure categorically and the subjective measure as a continuous variable. However, when measured in this manner the results were similar: Upward comparison had a positive effect on self-concept with the objective measure and a negative effect with the subjective measure.
Perhaps, then, the differential could be explained by the test scores themselves? Measurement error in the test scores could mean that the person choosing a comparison target was really more intelligent, or that the comparison target was really less intelligent, than the test scores would imply. Previous research suggests that individuals tend to choose comparison targets who are similar to them (e.g., Nosanchuk & Erickson, 1985; Wheeler, 1966). In this regard, the achievement levels of the comparison target could be considered as one indicator of the achievement level of the student. What does it mean when students choose a more able target person? The typical interpretation is that the student is upwardly comparing. However, it might also mean that the student actually is more intelligent than the test score (which has a substantial component of measurement error) would imply. According to this explanation, students who select more intelligent target comparisons are more intelligent students and, thus, should have higher academic self-concepts than reflected by their test scores. In this sense, the comparison target’s achievement level is like another achievement test itself. Which is the better explanation? One way is to ask the students whether the comparison person is more intelligent or not – the subjective measure of comparison direction. In contrast to the objective measure, these results suggest that upward comparison in the choice of comparison target has a negative effect on self-concept, results that are consistent with the BFLPE. Beyond the scope of the current study, but an important area for future research, is an examination of the reliability of these test scores (test scores which were based on standardised tests administered by the French Government). However, when the focus is on self-concept, as was the case for the current study, how one views a comparison subjectively is a more important consideration than an objective measure of the comparison. After all, how one perceives oneself in relation to one’s comparison target is more critical to one’s self-concept than any objective measure.

**Individual Ability as a Moderator of the BFLPE**

In the current study, compared to high-ability students, larger BFLPEs were associated with students of lower ability, although this pattern of results was more pronounced for French than math. That the BFLPE was more pronounced for low-
ability students can be explained in social comparison terms. Compared to low-ability settings, in high-ability classes there would have been many more opportunities for these low-ability students to compare themselves with extremely high-achieving peers. Additionally, in high-ability classes the average achievement levels tend to be higher than those of low-ability environments. So, for a low-ability student in a low-ability class, the average achievement level may not be completely unattainable. However, for a student of a similar low-ability level in a high-ability class, the average achievement level may seem unattainable. Both of these comparisons, with extremely high-achieving peers and with the average ability level of the class, may have contributed to low-ability students suffering more from the BFLPE than high-ability students. Moreover, high-ability students in high-ability settings were not as affected by the BFLPE as low-ability students. Presumably, these students were able to make self-enhancing downward comparisons, thereby protecting and bolstering their self-esteem (Wood, 1989). This view is consistent with research that demonstrates that downward comparisons make one feel better about oneself (e.g., Morse & Gergen, 1970; VanderZee et al., 1996; Wheeler & Miyake, 1992; Wills, 1981).

Although individual ability was assessed in both previous studies of the current thesis and has been examined in many previous studies, results have been inconsistent (e.g., Coleman & Fults, 1985; Marsh et al., 1995; Marsh & Hau, 2003; Marsh et al., 2000; Marsh & Rowe, 1996; Reuman, 1989), and the general consensus has been that the BFLPE generalises well over different levels of ability. That students of high-ability suffered less from the BFLPE, as demonstrated in the current study, only adds to the inconsistencies already apparent in the literature. Consequently, future research should concentrate on discovering the underlying causes of these inconsistencies.

**Achievement Goal Orientation as a Moderator of the BFLPE**

The only achievement goal orientation to moderate the BFLPE was a performance avoidance goal orientation, and then only for French. In high-ability settings students who held more of a performance avoidance orientation had lower French self-concepts than those who were low in performance avoidance. Hence, the
BFLPE was more evident for those who held a higher performance avoidance goal orientation. Perhaps the BFLPE was more pronounced for these students due to the anxiety underlying this goal orientation. Performance avoidance individuals are afraid of performing poorly. Being in a high-ability situation may have made these students even more anxious about their performance, resulting in their self-concepts being especially vulnerable. Although this scale was unreliable and thus these results should be considered cautiously, one could argue that with a more reliable scale the effect may be larger. Future research is necessary to test this proposition. For math, none of the achievement goal orientations moderated the BFLPE. Irrespective of the achievement goal orientation held, students in high-ability classes had lower math self-concepts than their counterparts in low-ability classes.

Consistent with implications drawn from previous research (Elliot & Church, 1997; Elliot & McGregor, 2001; Elliot et al., 2005), whereas mastery and performance approach goal orientations had a positive association with self-concept for both French and math, a mastery avoidance goal orientation had a negative association with academic self-concept. Mastery goals have been related to intrinsic motivation. Students who hold a mastery approach goal orientation strive to learn, develop, and improve their skills. Perhaps students who held this type of goal orientation had higher academic self-concepts because they were motivated to learn and focussed on maximising their learning potential. Performance approach goals have been associated with higher achievement. Hence, students who espoused these goals may have taken comfort in their accomplishments, with the resulting positive effect on academic self-concept. Students who held a mastery avoidance goal orientation had lower academic self-concepts. The overriding concern for these students is that they may fail to learn or understand all the necessary material. This uncertainty may have created anxiety about their abilities that translated into lower academic self-concepts.

**Strengths and Limitations**

The major strength of the current study was in its use of standardised achievement tests and psychometrically sound measurement instruments. Using standardised tests provided a measure of academic achievement that was comparable
across all students in the study. Additionally, the use of the ASDQ II, a psychometrically sound measure of academic self-concept, provided a more reliable self-evaluation measure than was available in Study 2. The AGQ also proved to be psychometrically sound with the exception of the performance avoidance scales. The use of these more appropriate instruments rendered the inferences drawn from the results of the current study more valid and reliable.

The present investigation also benefited from the use of state-of-the-art statistical techniques. Multilevel modelling is a necessary analytical tool to use when data are nested as they are in the current sample. Students in these data were grouped together in classes and classes were grouped together in schools. As individuals in one group (e.g., a class or school) are more alike than they are to individuals from other groups, ignoring a nested structure such as this can lead to serious statistical problems. Multilevel modelling overcomes these problems by partitioning the variance from each group level.

As was the case with Studies 1 and 2, the present investigation also used correlational data, so no causality can be inferred. Hence, for example, it is unknown whether holding a mastery approach goal orientation causes higher academic self-concept, or whether having a high academic self-concept causes one to adopt a mastery approach goal orientation. This is a major issue that future research needs to address. This limitation could be overcome by conducting longitudinal causal modelling studies that measure variables before and after students enter selective education. In this way, the determinants and consequences of high-ability environments on self-concept and other psychosocial drivers of life potential could be elucidated.

A further limitation of the current study was that not all students reported a comparison target and so multiple imputation was used to replace these missing data. Missing data are an enormous problem, not only for all questionnaire type social comparison research, but also for survey research in general. The present study addressed this issue by conducting analyses that demonstrated similar results to those of the imputed data for those for whom comparison information was unavailable. In doing so, the use of multiple imputation was considered to have been an appropriate strategy to overcome problems associated with missing data. Although these missing
data may reflect actual practice (students genuinely did not compare), it may also be that students did compare but were unable to consciously articulate exactly with whom they compared themselves. It also might be plausible that students do not compare themselves with just one individual but with a sub-group of individuals. These suggestions should be elucidated by further research. For example, there is growing recognition that quantitative and qualitative research methods complement each other in ways that both consolidate findings and shed light on issues that are best studied by one or the other (e.g., Martin, Marsh, Williamson, & Debus, 2003). Hence, qualitative research that addresses whether or not students compare, how they use the comparisons if they do compare, and how they feel about these comparisons, would be a useful adjunct to quantitative research.

**Implications for Theory, Research, and Practice**

Findings from the current study present further support for implications proposed in Study 2 concerning BFLPE and social comparison theories. The current findings demonstrated that the BFLPE could be moderated by selected individual comparisons, measured objectively, and suggest that comparing upward may reduce the BFLPE. However, this moderating effect, a moderating effect of an objective comparison on the BFLPE for French, was one of only two times in this thesis that such an effect was found. In addition, the current moderating effect in French should be considered in light of the results for the objective comparison measure for math and the subjective comparison measure that did not moderate the BFLPE. When measures of academic self-concept, like that used in the current study, include items that tap into how one compares to others, the way one subjectively perceives oneself in regard to a comparison target may be a more important consideration for determining self-concept than an objective measure of one’s abilities. Additionally, as previously discussed, what is perceived as an upward comparison, when measured objectively, may actually be a product of measurement error in the tests. Hence, pending further research, the current findings suggest that the BFLPE is not moderated by selected comparisons with an individual.

Furthermore, the two processes – comparing with a selected individual and the BFLPE – appear to co-exist. To the extent that they do, it could be argued that forced
comparisons with a generalised other may be more important in determining the BFLPE than selected comparisons alone. Alternatively, the BFLPE may be the result of a combination of selected individual comparisons and forced generalised comparisons. However, further research is necessary to elucidate the contributions of each to the BFLPE. Hence, as noted in Chapter 5, in order to further inform both BFLPE and social comparison theory, further research is necessary to investigate the nature and mechanisms of forced comparisons with a generalised other, perhaps using a longitudinal causal modelling design. An especially fruitful area may be to discover the cognitive processes that students use when making forced comparisons and when comparing with a generalised other.

The current findings also have implications for future research. For example, the current study was a snapshot of a single timeframe. Comparison processes may change over time and the way they relate to the BFLPE may also change. A longitudinal study following students over the course of several years could ascertain if such changes exist and also examine the causality of effects. Differences in the way individuals use social comparisons may also exist between students in western and non-western countries. Cross-cultural research combining comparison information with the BFLPE may help to establish whether such differences exist. Furthermore, the differential results noted in the current study for French and math should be investigated to discover the underlying causes of such differences.

An additional avenue for future research is to examine class- and school-average ability simultaneously. Studies assessing the BFLPE generally rely on class-average ability or school-average ability to determine the size of the BFLPE. Very few studies have examined the effect on academic self-concept of placing both these measures in the same model simultaneously in order to ascertain whether one is more important than the other in the formation of the BFLPE, or if they both contribute equally. Although the focus of the current study was on class-average ability, school-average ability data were available. In unreported analyses, school-average ability was added to each of the models tested. In most cases, class-average ability was the determining factor of the BFLPE. It is not really surprising that class-average ability was more important than school-average ability in determining the BFLPE. Students in the current study spent the entire year in the same class with the same group of
students, and so it is likely that they would have been aware of each other’s relative performance. In contrast, especially as these were students in their first year of high school, students may not have been as aware of the performances of students in other classes within the school. However, while it is true of this study, class-average ability may not always be more important in determining the BFLPE. For example, students in their final years of high school may be more aware of their school’s average ability in relation to other schools in their area. Future research could investigate the circumstances under which class-average ability or school-average ability becomes more important in determining the BFLPE.

The current study provides evidence that the achievement goal one espouses can affect one’s self-concept, and consequently can offer suggestions for practice. It appears that holding an approach goal orientation, be that mastery or performance, can have a positive effect on academic self-concept and so students should be encouraged to adopt such a motivational style. On the contrary, adopting a mastery avoidance goal orientation can have a deleterious effect on academic self-concept, and should be discouraged. Rather, an effort should be made to educate mastery avoidance individuals on the use of approach goals. Moreover, the negative effect on academic self-concept is even more exacerbated for students who hold a performance avoidance goal orientation when they attend high-ability schools and classes. Students who hold a performance avoidance goal orientation are afraid of performing poorly and attending a high-ability school may exacerbate this fear, with a resulting drop in self-concept. Thus, an important issue for high-ability schools to address is the early detection of these students with the intent of teaching them more positive motivational styles and ways of dealing with their fears and anxiety, in order to bolster their self-concepts.

Furthermore, these findings underscore evidence provided by Marsh and his colleagues concerning the negative effects of attending high-ability schools and have implications for educational policy makers worldwide. Although academically selective education is a very popular and much sought after educational strategy for nurturing our most intelligent students, these results suggest that segregation on the basis of academic ability is detrimental to a student’s self-worth. Moreover, as lower academic self-concept has been associated with many negative educational outcomes
that have been shown to be long lasting, these results suggest that academic segregation may not be the optimal environment for high-ability students to realise their full potential and imply that it is necessary to put strategies in place in selective environments to limit the BFLPE.

**Summary**

This chapter presented the background, aims, hypotheses, research questions, rationales for the hypotheses and research questions, and results for Study 3. Selected social comparisons had a differential effect on the BFLPE depending on how they were measured. When measured objectively, the BFLPE was moderated, but for French only: The BFLPE was more evident for students in high-ability classes who compared with low-performing others. When measured subjectively, selected social comparisons did not moderate the BFLPE. However, the BFLPE was moderated by individual ability and a performance avoidance goal orientation. The BFLPE was more pronounced for students of lower ability and for those who held a higher performance avoidance goal orientation. The chapter concluded with a discussion of the findings, focussing on the differential between results for the objective and subjective measures of social comparison in relation to the BFLPE. The strengths and limitations of the current study and implications of the findings were also discussed. Overall, this study contributed to the BFLPE and social comparison literatures by demonstrating that selected comparisons with individuals co-existed with the BFLPE.
CHAPTER 7

SUMMARY AND CONCLUSIONS

The focus of the present investigation was to extend current BFLPE theory and research in order to elucidate potentially potent processes and constructs to inform new solutions for enabling high-ability students to reach their full potential. More specifically, the broad aims of the present investigation were to: (a) test the external validity of the BFLPE across 41 countries in order to ascertain whether support can be identified for the universality of this theory; (b) test whether the BFLPE is evident in economically developing countries and also in collectivist countries in order to ascertain whether the BFLPE transcends cultural and economic barriers; (c) investigate potential moderators of the BFLPE to identify constructs that may moderate the adverse effects of the BFLPE; (d) elucidate whether upward social comparisons moderate or co-exist with the BFLPE to resolve a conflict in the literature; (e) critically analyse the relation between social comparison processes and the BFLPE to further inform theory; and (f) test whether the BFLPE varies as a function of ability to clarify the impact of the BFLPE on different levels of student ability. Three synergistic studies, capitalising on cutting-edge advances in statistical methodology and BFLPE theory and research, were conducted to achieve these broad aims.

Concerning the first aim, Study 1 assessed the external validity of the BFLPE. Findings demonstrated that the BFLPE was evident in a cross-national sample of 41 culturally and economically diverse countries, and individually in 38 of the 41 countries. Importantly, the effect of school-average ability (the BFLPE) was negative in all 41 individual countries tested. Furthermore, in relation to the second aim, the three countries where this effect was not significant comprised individualist, collectivist and developed countries. By demonstrating the existence of the BFLPE in such a culturally diverse sample, Study 1 provided evidence of the external validity and universal applicability of the BFLPE and established the BFLPE as a truly universally applicable theory. Consequently, Study 1 demonstrated that the
BFLPE applies to countries worldwide, irrespective of their cultural orientation or economic prosperity.

As regards the third aim, Study 1 conducted a search for potential BFLPE moderators. Small moderating effects for the BFLPE were found for SES, individual differences in learning, and individual perceptions of the learning environment. Results for SES suggested that the BFLPE – the negative effect of school-average ability – was marginally smaller for students in high-ability schools when a parent was employed in a more prestigious occupation or when there were more educational resources available at home. Moreover, slightly smaller BFLPEs were associated with students who: (a) used elaboration as a learning strategy; (b) were motivated either extrinsically or intrinsically; (c) felt able to succeed; (d) had more positive attitudes to school; or (e) felt a sense of belonging to the school. However, these interaction effects were trivial and in a more moderately sized sample none of these effects may have been statistically significant. Larger interaction effects were found for constructs that were associated with more pronounced BFLPEs. Larger BFLPEs were associated with students who: (a) used memorisation as a learning technique; (b) preferred to work in a cooperative environment; or (c) were anxious, although the size of the first two effects was small considering the size of the sample (in excess of a quarter of a million students) and the power of the tests. In addition, Study 3 examined the moderating effects of achievement goal orientations on the BFLPE. The only achievement goal orientation found to moderate the BFLPE was a performance avoidance goal orientation, and then only for one academic subject. Nevertheless, these results do offer potential for intervention strategies. For example, interventions that focus on reducing anxiety levels in susceptible students in high-ability environments may help to counteract the BFLPE. Promoting a sense of belonging in high-ability classes and schools and encouraging more cooperative learning environments may also be beneficial, but further research is necessary to fully elucidate these moderators.

In relation to the fourth aim, Study 2 tested whether selected social comparisons impacted the BFLPE by investigating an apparent discrepancy between the BFLPE results of Marsh and his colleagues in relation to social comparison and academic self-concept, and the social comparison findings of Blanton et al. (1999) and Huguet.
et al. (2001) in relation to actual performance. BFLPE theory argues that lower academic self-concepts are the result of comparisons with classmates and lower academic self-concepts have been associated with poorer performance (Marsh & Craven, 2005, 2006; Marsh & Yeung, 1997b; Valentine & DuBois, 2005). However, two social comparison studies in particular (Blanton et al., 1999; Huguet et al., 2001) demonstrated that upward comparisons improved performance but had no effect on self-evaluation. In order to examine this contradiction, Study 2 further analysed the social comparison studies of Blanton et al. and Huguet et al. to identify whether the negative effects of the BFLPE on self-concept co-existed with, or were moderated by, the positive effects of upward comparison on performance. With one exception (Dutch), results from Study 2 indicated that upward social comparisons did not moderate the BFLPE. Rather the BFLPE co-existed with selected upward comparisons that enhanced performance. Evidently, choosing to compare with a more able individual student enhances performance, but at the same time makes one feel distinctly inferior about one’s abilities.

As Study 2 was conducted to assess the effect of social comparisons on performance, it was not designed to investigate the BFLPE. Using instrumentation more appropriate for examining the BFLPE, Study 3 addressed the fifth aim by elucidating the relation between the BFLPE and selected social comparisons. When measured objectively, the BFLPE was only moderated by selected social comparisons in one instance (French). When measured subjectively, the BFLPE was not moderated by social comparisons that individuals selected. Rather than interacting with each other, the BFLPE and the effects of selected social comparisons on self-concept co-existed. Results from Study 3 also demonstrated that the main effect of selected comparisons had a differential effect on academic self-concept depending on whether they were measured objectively or subjectively. When measured objectively, selected upward comparisons had a positive effect on academic self-concept. However, when measured subjectively, selected upward comparisons decreased academic self-concept and selected downward comparisons increased it. Possible reasons for this differential in the findings were proposed in Chapter 6.
A number of commonalities were either studied or emerged across these studies. In relation to the sixth aim, individual ability, the BFLPE moderator that has received most attention in previous BFLPE studies was investigated in all three studies of the present investigation. In Study 1 there was a significant linear interaction of individual ability with school-average ability, such that the BFLPE was somewhat greater for high-ability students attending high-ability schools. However, due to the large size of the sample in that study and the small size of the interaction effect, the conclusion was that the BFLPE generalised well across all levels of ability. In Study 2, individual ability moderated the BFLPE for one academic subject (math) out of the 10 subjects tested. Once again, the BFLPE was more pronounced for higher-achieving students than lower-achieving students in high-ability classes. Nonetheless, the size of this interaction was small, and given that there was a significant interaction for only one academic subject, it was again concluded that the BFLPE was consistent across ability levels. Study 3 produced results inconsistent with those of the previous two studies. In this study it was students of lower ability who suffered the negative effects of the BFLPE to a greater extent than students of higher ability.

The differences between the three studies in results for the moderating effect of individual ability could be accounted for by the different measures of academic self-concept and ability used in each of the studies. In Study 1, five items were used to measure math self-concept and in Study 3 the ASDQ II, based on six items, was used. Although the items used in these two studies were similar, they were not identical. In Study 2, a single-item measure of self-evaluation in particular subjects was used as a measure of self-concept and hence was highly dissimilar to those used in Studies 1 and 2. In addition, the observed items on which the measures of academic self-concept were based were measured using different scales. Furthermore, individual ability, and by extension, class- and school-average ability was measured differently across the three studies. In Study 1, individual ability was estimated from a student’s test score using plausible values; in Study 2 teacher-assigned grades were used; and Study 3 used standardised test scores. As different measures of academic self-concept and individual ability were used across the three studies of the present investigation, it is unclear whether similar constructs were in fact being measured in each of the studies. As such, it is perhaps not surprising that
results were conflicting. However, the inconsistencies noted between measures used and results found in the studies of the present investigation are reflective of the literature in this area in general. Not only are different measures typically used, but interactions between class- and school-average ability and individual student ability levels are usually small and inconsistent in direction. Consequently, future research should concentrate on discovering the underlying causes of these inconsistencies. Furthermore, as a result of the inconsistencies in the current findings, one must agree with Marsh and Craven (2002) that the BFLPE is reasonably invariant across ability levels.

Comparisons that individuals are free to select, and their moderating effect on the BFLPE, were another common thread across the studies. These comparisons were the focus of Studies 2 and 3. Except for results concerning the language of instruction, the BFLPE was not moderated by comparisons that individuals selected. Rather, results suggest that the two processes – comparing with a selected individual and the BFLPE – co-exist. Study 2 demonstrated that the BFLPE co-existed with the positive effects of upward comparisons on performance. In Study 3, the BFLPE co-existed with the effects of selected social comparisons on academic self-concept when measured subjectively and when measured objectively for math. (Differences in results between moderating results for French and math and between subjective and objective measures of social comparison were discussed in Chapter 6). These results have important implications for BFLPE and social comparison theories (see Chapters 5 and 6). They suggest that selected social comparisons with an individual may not be solely responsible for the BFLPE. Rather, the BFLPE may be the product of forced comparisons with a generalised other, or the combination of selected individual comparisons with forced generalised comparisons may be responsible for the effect. Further research is required to explicate the mechanisms underlying the BFLPE and in particular to establish the form and nature of the generalised other and the processes by which forced comparisons with the generalised other and selected individual comparisons may operate together in the formation of the BFLPE.

The impact of anxiety on the BFLPE was a common theme that emerged across two of the studies. To date research has had limited success in investigating factors that moderate the BFLPE. Hence, to find that anxiety plays a role in the BFLPE is a
major finding of the present investigation and has important implications for practice, as noted in Chapter 4. Study 1 revealed that highly anxious students were more affected by the BFLPE. Whereas highly anxious students in high-ability schools experienced a large drop in math self-concept, the decline for less anxious students in high-ability schools was minimal. Anxiety was not directly measured in Study 3. However, in that study the BFLPE was moderated by a performance avoidance goal orientation, whereby students in high-ability classes who held a high performance avoidance goal orientation in French suffered the BFLPE to a greater extent than those who held a low performance avoidance goal orientation. Individuals high in performance avoidance have an underlying performance anxiety, as they are fearful of performing poorly. Attending high-ability classes may have made these students even more anxious about their performance, with a corresponding drop in their self-concepts, thus contributing to the BFLPE. These latter results, however, need to be considered cautiously and future research could benefit from using sound psychometric measures of performance avoidance.

Consequently, these findings suggest that anxiety is an important facet of the BFLPE, such that highly anxious students are more likely to suffer its negative effects than students with low anxiety levels. As noted in Chapter 4, it is unclear whether students bring their anxieties with them to high-ability schools and classes, or whether they become anxious as a result of attending such environments; nonetheless, it is evident that reducing anxiety levels in students prone to this condition would be an important step forward in overcoming the BFLPE.

Discussion of findings for each study included an analysis of the strengths and limitations for that particular study. However, these can also be found across all studies. For example, the three studies of the present investigation capitalised on advances in statistical analyses by utilising multilevel modelling to analyse data in all three studies. When data are comprised of different levels nested within each other, such as students within classes within schools, it is necessary to use statistical methods that take these levels into account when performing calculations (see Chapter 4). As every study of the present investigation contained a multilevel structure, multilevel modelling was used in order to avoid serious statistical problems associated with using single level techniques to analyse multilevel data. As such, the
use of multilevel modelling for data analysis is a significant strength of the present investigation. Another strength of the present investigation lies in its use of large samples in all three studies. When multilevel modelling is used, larger samples are preferred to ensure that sample sizes at all levels (e.g., number of classes) are adequate. Furthermore, the large sample in Study 1 allowed the external validity of the BFLPE to be tested in ways that otherwise would not have been possible with a smaller sample. A limitation of the present investigation that applies across the three studies is that different measures of self-concept and ability were used in each study. However, as the three studies were undertaken for different reasons and under different circumstances, it is not surprising that these measures were not similar. Although it may not always be possible to use similar measures of ability, it is, however, important that future research, especially research examining the link between social comparison and the BFLPE or tests of the moderating effect of individual ability on the BFLPE, uses similar measures of self-concept.

The present investigation has advanced current knowledge of the BFLPE from both a theoretical and a practical perspective. Theoretically, it has established the BFLPE as a universally applicable theory by demonstrating that it applies to students everywhere, irrespective of their nation’s culture or economic development. Furthermore, the present investigation has begun the process of elucidating the psychological processes thought to underlie the BFLPE. Results have indicated that the comparisons that individuals select co-exist with, and do not moderate, the BFLPE.

From a practical perspective, the present investigation identified many constructs that alleviated or reduced the BFLPE. Although implications arising from many of these constructs could be considered as suggestive only, one construct, anxiety, has important practical applications. Reducing anxiety levels for highly anxious students who attend high-ability classes and schools is an important avenue to pursue in future intervention strategies in order to alleviate the negative effects associated with the BFLPE, thus allowing high-ability students to benefit from selective education and reach their full potential.
LIST OF APPENDICES

Appendix A
Calculation of Standard Errors........................................352

Appendix B
ICC Calculations.........................................................354

Appendix C
Multiple Imputation and Questionnaire for Study 3...............369
APPENDIX A

This Appendix provides details of the steps taken to average parameter estimates and calculate accurate standard errors for plausible values in Study 1 and the imputed data sets in Study 3. These steps are those recommended in the PISA documentation (OECD, 2005a) and are taken from Raudenbush, Bryk, and Congdon (2005). These calculations resulted in, not only averaged parameter estimates, but also standard errors that reflected variance both within and between the plausible values and imputed data sets.

Study 1

1. The five parameter estimates were averaged:

\[ t = \frac{\text{sum of parameter estimates of the 5 plausible values}}{5} \]

\[ t = \frac{\sum_{m=1}^{M} t_m}{M} \]

2. The variances of the parameters from the five analyses were averaged:

\[ U = \frac{\text{sum of the variances of plausible values 1-5}}{5} \]

\[ U = \frac{\sum_{m=1}^{M} U_m}{M} \]

3. The variance of each of the parameters from the average parameter was estimated:

\[ B = \frac{\text{parameter estimates for plausible values 1-5 minus the average (calculated in step 1), squared and then summed}}{4} \]

\[ B_m = \frac{\sum_{m=1}^{M} (t_m - t)^2}{(M - 1)} \]

4. The outcomes of steps 2 and 3 were summed:

\[ V = U + (1.2)B \]

\[ V = U + (1 + M^{-1})B_m \]
Study 3

1. The five parameter estimates were averaged:

\[ t = \text{sum of parameter estimates from 5 data sets} \]

\[ t = \frac{\sum_{m=1}^{M} t_m}{M} \]

2. The variances of the parameters from the five analyses were averaged:

\[ U = \text{sum of the variances of parameter estimates from 5 data sets} \]

\[ U = \frac{\sum_{m=1}^{M} U_m}{M} \]

3. The variance of each of the parameters from the average parameter was estimated:

\[ B = \text{parameter estimates from 5 data sets minus the average (calculated in step1), squared and then summed} \]

\[ B_m = \frac{\sum_{m=1}^{M} (t_m - t)^2}{(M - 1)} \]

4. The outcomes of steps 2 and 3 were summed:

\[ V = U + (1.2)B \]

\[ V = U + (1+M^{-1})B_m \]
APPENDIX B

This appendix provides details of the calculations (as noted in Study 2) for:
1. The ICC for the Netherlands using the PISA (2003) data (school level)
2. The ICC for France using the PISA (2003) data (school level)
3. The ICC for the subjects in the Blanton et al. (1999) study using modified grades (class and school levels)
4. The ICC for the subjects in the Blanton et al. (1999) study using unmodified grades (class and school levels)
5. The ICC for the subjects in the Huguet et al. expanded data at both the class and school levels.

1. Calculating ICC for the Netherlands using PISA 2003 Data
   (Study 2a)

\[ z_{pi1math} \sim \mathcal{N}(\mu_{pi}, \Omega) \]
\[ z_{pi1math} = \beta_{0i} + \Omega \]
\[ \beta_{0i} = 0.569(0.054) + \mu_{0i} + \varepsilon_{0i} \]

\[ \begin{bmatrix} \mu_{0i} \\ \varepsilon_{0i} \end{bmatrix} \sim \mathcal{N}(0, \Omega_{0i}) \quad : \quad \Omega_{0i} = \begin{bmatrix} 0.431(0.033) \end{bmatrix} \]

\[ \begin{bmatrix} \varepsilon_{0i} \end{bmatrix} \sim \mathcal{N}(0, \Omega_{0i}) \quad : \quad \Omega_{0i} = \begin{bmatrix} 0.296(0.008) \end{bmatrix} \]

\[-2 \cdot \text{loglikelihood(IGLS Deviance)} = 6799.481, \text{of 3860 cases in use}\]
\[ z_{pv2math_{ij}} \sim N(\lambda, \Omega) \]
\[ z_{pv2math_{ij}} = \beta_{0j} \text{CONS} \]
\[ \beta_{0j} = 0.565(0.055) + \mu_{0j} + e_{0j} \]
\[
\begin{bmatrix}
\mu_{0j} \\
e_{0j}
\end{bmatrix} \sim N(0, \Omega_0) : \Omega_0 = \begin{bmatrix} 0.434(0.033) \end{bmatrix}
\]
\[
\begin{bmatrix}
\varepsilon_{0ij}
\end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.303(0.008) \end{bmatrix}
\]

\[-2*\text{loglikelihood(IGLS Deviance)} = 6888.180 (3860 of 3860 cases in use)\]

\[ z_{pv3math_{ij}} \sim N(\lambda, \Omega) \]
\[ z_{pv3math_{ij}} = \beta_{0j} \text{CONS} \]
\[ \beta_{0j} = 0.564(0.054) + \mu_{0j} + e_{0ij} \]
\[
\begin{bmatrix}
\mu_{0j} \\
e_{0ij}
\end{bmatrix} \sim N(0, \Omega_0) : \Omega_0 = \begin{bmatrix} 0.427(0.032) \end{bmatrix}
\]
\[
\begin{bmatrix}
\varepsilon_{0ij}
\end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.302(0.008) \end{bmatrix}
\]

\[-2*\text{loglikelihood(IGLS Deviance)} = 6873.226 (3860 of 3860 cases in use)\]

\[ z_{pv4math_{ij}} \sim N(\lambda, \Omega) \]
\[ z_{pv4math_{ij}} = \beta_{0j} \text{CONS} \]
\[ \beta_{0j} = 0.557(0.054) + \mu_{0j} + e_{0j} \]
\[
\begin{bmatrix}
\mu_{0j} \\
e_{0j}
\end{bmatrix} \sim N(0, \Omega_0) : \Omega_0 = \begin{bmatrix} 0.430(0.033) \end{bmatrix}
\]
\[
\begin{bmatrix}
\varepsilon_{0ij}
\end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.300(0.008) \end{bmatrix}
\]

\[-2*\text{loglikelihood(IGLS Deviance)} = 6857.536 (3860 of 3860 cases in use)\]
$z_{pv5math_i} \sim \mathcal{N}(XB, \Omega)$
$z_{pv5math_i} = \beta_{0i} \text{CONS}$
$\beta_{0i} = 0.568(0.054) + \mu_{0i} + e_{0i}$

\[
\begin{bmatrix}
\mu_{0i} \\
\end{bmatrix} \sim \mathcal{N}(0, \Omega_{\mu}) : \Omega_{\mu} = \begin{bmatrix} 0.426(0.032) \end{bmatrix}
\]

\[
\begin{bmatrix}
e_{0i} \\
\end{bmatrix} \sim \mathcal{N}(0, \Omega_{\varepsilon}) : \Omega_{\varepsilon} = \begin{bmatrix} 0.297(0.008) \end{bmatrix}
\]

$-2 \cdot \text{loglikelihood(IGLS Deviance)} = 6812.997 (3860 \text{ of } 3860 \text{ cases in use})$

Averaged across all plausible values:
$U_{0j} = .4296$
$e_{01j} = .2996$

$ICC = \frac{.4296}{.4296 + .2996} = .5891$
2. Calculating ICC for France using Pisa 2003 Data (Study 2b)

\[
z_{p1\text{math}}_j \sim N(XB, \Omega)
\]
\[
z_{p1\text{math}}_j = \beta_{0j} \text{CONS}
\]
\[
\beta_{0j} = 0.263(0.047) + u_{0j} + e_{0j}
\]

\[
\begin{bmatrix}
u_{0j} \\
e_{0ij}
\end{bmatrix} \sim N(0, \Omega_0) : \Omega_0 = \begin{bmatrix} 0.342(0.029) \end{bmatrix}
\]

\[
\begin{bmatrix}
u_{0j} \\
e_{0ij}
\end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.398(0.010) \end{bmatrix}
\]

-2*loglikelihood(IGLS Deviance) = 8461.562 (4154 of 4154 cases in use)

\[
z_{p2\text{math}}_j \sim N(XB, \Omega)
\]
\[
z_{p2\text{math}}_j = \beta_{0j} \text{CONS}
\]
\[
\beta_{0j} = 0.258(0.047) + u_{0j} + e_{0j}
\]

\[
\begin{bmatrix}
u_{0j} \\
e_{0ij}
\end{bmatrix} \sim N(0, \Omega_0) : \Omega_0 = \begin{bmatrix} 0.337(0.029) \end{bmatrix}
\]

\[
\begin{bmatrix}
u_{0j} \\
e_{0ij}
\end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.402(0.010) \end{bmatrix}
\]

-2*loglikelihood(IGLS Deviance) = 8498.521 (4154 of 4154 cases in use)

\[
z_{p3\text{math}}_j \sim N(XB, \Omega)
\]
\[
z_{p3\text{math}}_j = \beta_{0j} \text{CONS}
\]
\[
\beta_{0j} = 0.257(0.047) + u_{0j} + e_{0j}
\]

\[
\begin{bmatrix}
u_{0j} \\
e_{0ij}
\end{bmatrix} \sim N(0, \Omega_0) : \Omega_0 = \begin{bmatrix} 0.338(0.029) \end{bmatrix}
\]

\[
\begin{bmatrix}
u_{0j} \\
e_{0ij}
\end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0.400(0.011) \end{bmatrix}
\]

-2*loglikelihood(IGLS Deviance) = 8482.460 (4154 of 4154 cases in use)
$z_{pV4math_{ij}} \sim N(\chi B, \Omega)$
$z_{pV4math_{ij}} = \beta_{0ij} \text{CONS}$
$\beta_{0ij} = 0.252(0.047) + \mu_{0ij} + \epsilon_{0ij}$

$\begin{bmatrix} \mu_{0ij} \end{bmatrix} \sim N(0, \Omega_\mu) : \Omega_\mu = \begin{bmatrix} 0.342(0.029) \end{bmatrix}$

$\begin{bmatrix} \epsilon_{0ij} \end{bmatrix} \sim N(0, \Omega_\epsilon) : \Omega_\epsilon = \begin{bmatrix} 0.394(0.010) \end{bmatrix}$

$-2^*\text{loglikelihood(IGLS Deviance)} = 8419.513$ (4154 of 4154 cases in use)

$z_{pV5math_{ij}} \sim N(\chi B, \Omega)$
$z_{pV5math_{ij}} = \beta_{0ij} \text{CONS}$
$\beta_{0ij} = 0.257(0.047) + \mu_{0ij} + \epsilon_{0ij}$

$\begin{bmatrix} \mu_{0ij} \end{bmatrix} \sim N(0, \Omega_\mu) : \Omega_\mu = \begin{bmatrix} 0.342(0.027) \end{bmatrix}$

$\begin{bmatrix} \epsilon_{0ij} \end{bmatrix} \sim N(0, \Omega_\epsilon) : \Omega_\epsilon = \begin{bmatrix} 0.411(0.011) \end{bmatrix}$

$-2^*\text{loglikelihood(IGLS Deviance)} = 8588.561$ (4154 of 4154 cases in use)

Averaged across all plausible values:

$U_{0j} = .3402$
$\epsilon_{0ij} = .401$

$\text{ICC} = \frac{.3402}{.3402 + .401} = .45898$
3. Calculating ICC for Blanton et al. (1999) for modified grades (Study 2a)

Geography

\[ Z_{ctvgeo_{ijk}} \sim N(XB, \Omega) \]
\[ Z_{ctvgeo_{ijk}} = \beta_{0ijk} \text{CONS} \]
\[ \beta_{0ijk} = 0.008(0.106) + \gamma_{0k} + \nu_{0jk} + \epsilon_{0ijk} \]

\[ [\gamma_{0k}] \sim N(0, \Omega_\gamma) : \Omega_\gamma = [0.000(0.000)] \]

\[ [\nu_{0jk}] \sim N(0, \Omega_\nu) : \Omega_\nu = [0.346(0.091)] \]

\[ [\epsilon_{0ijk}] \sim N(0, \Omega_\epsilon) : \Omega_\epsilon = [0.646(0.031)] \]

\[-2*\text{loglikelihood(TIGLS Deviance)} = 2283.054 (913 of 920 cases in use)\]

\[ ICC = \frac{.346}{.346 + .646} = .348 \]

Biology

\[ Z_{ctvbio_{ijk}} \sim N(XB, \Omega) \]
\[ Z_{ctvbio_{ijk}} = \beta_{0ijk} \text{CONS} \]
\[ \beta_{0ijk} = 0.007(0.106) + \gamma_{0k} + \nu_{0jk} + \epsilon_{0ijk} \]

\[ [\gamma_{0k}] \sim N(0, \Omega_\gamma) : \Omega_\gamma = [0.000(0.000)] \]

\[ [\nu_{0jk}] \sim N(0, \Omega_\nu) : \Omega_\nu = [0.346(0.091)] \]

\[ [\epsilon_{0ijk}] \sim N(0, \Omega_\epsilon) : \Omega_\epsilon = [0.646(0.031)] \]

\[-2*\text{loglikelihood(TIGLS Deviance)} = 2280.702 (912 of 920 cases in use)\]

\[ ICC = \frac{.346}{.346 + .646} = .348 \]
English

$Z_{ctveng2_{ijk}} \sim N(X_{i}, \Omega)$

$Z_{ctveng2_{ijk}} = \beta_{0ijk} \text{CONS}$

$\beta_{0ijk} = 0.008(0.106) + \nu_{0k} + \nu_{0jk} + \varepsilon_{0ijk}$

$\nu_{0k} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.000(0.000) \end{bmatrix}$

$\nu_{0jk} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.346(0.091) \end{bmatrix}$

$\varepsilon_{0ijk} \sim N(0, \Omega_{\varepsilon}) : \Omega_{\varepsilon} = \begin{bmatrix} 0.646(0.031) \end{bmatrix}$

$-2*\text{loglikelihood(IQLS Deviance)} = 2278.201 (911 \text{ of } 920 \text{ cases in use})$

$ICC = \frac{.346}{.346 + .646} = .348$

French

$Z_{ctvfre2_{ijk}} \sim N(X_{i}, \Omega)$

$Z_{ctvfre2_{ijk}} = \beta_{0ijk} \text{CONS}$

$\beta_{0ijk} = 0.005(0.106) + \nu_{0k} + \nu_{0jk} + \varepsilon_{0ijk}$

$\nu_{0k} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.000(0.000) \end{bmatrix}$

$\nu_{0jk} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.347(0.091) \end{bmatrix}$

$\varepsilon_{0ijk} \sim N(0, \Omega_{\varepsilon}) : \Omega_{\varepsilon} = \begin{bmatrix} 0.648(0.031) \end{bmatrix}$

$-2*\text{loglikelihood(IQLS Deviance)} = 2278.342 (910 \text{ of } 920 \text{ cases in use})$

$ICC = \frac{.346}{.346 + .646} = .348$
Dutch
\[ Z_{ctvdt}^{2}_{ijk} \sim N(XB, \Omega) \]
\[ Z_{ctvdt}^{2}_{ijk} = \beta_{0ijk} \text{CONS} \]
\[ \beta_{0ijk} = 0.008(0.106) + \nu_{0k} + \lambda_{ijk} + \epsilon_{0ijk} \]

\[
\begin{bmatrix}
\nu_{0k}
\end{bmatrix}
\sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.000(0.000) \end{bmatrix}
\]

\[
\begin{bmatrix}
\lambda_{ijk}
\end{bmatrix}
\sim N(0, \Omega_{\lambda}) : \Omega_{\lambda} = \begin{bmatrix} 0.346(0.091) \end{bmatrix}
\]

\[
\begin{bmatrix}
\epsilon_{0ijk}
\end{bmatrix}
\sim N(0, \Omega_{\epsilon}) : \Omega_{\epsilon} = \begin{bmatrix} 0.646(0.031) \end{bmatrix}
\]

\[-2* \text{loglikelihood(IGLS Deviance)} = 2283.054 (913 of 920 cases in use)\]

\[
\text{ICC} = \frac{.346}{.346 + .646} = .348
\]

Math
\[ Z_{ctvmth}^{2}_{ijk} \sim N(XB, \Omega) \]
\[ Z_{ctvmth}^{2}_{ijk} = \beta_{0ijk} \text{CONS} \]
\[ \beta_{0ijk} = 0.006(0.108) + \nu_{0k} + \lambda_{ijk} + \epsilon_{0ijk} \]

\[
\begin{bmatrix}
\nu_{0k}
\end{bmatrix}
\sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.000(0.000) \end{bmatrix}
\]

\[
\begin{bmatrix}
\lambda_{ijk}
\end{bmatrix}
\sim N(0, \Omega_{\lambda}) : \Omega_{\lambda} = \begin{bmatrix} 0.348(0.093) \end{bmatrix}
\]

\[
\begin{bmatrix}
\epsilon_{0ijk}
\end{bmatrix}
\sim N(0, \Omega_{\epsilon}) : \Omega_{\epsilon} = \begin{bmatrix} 0.644(0.031) \end{bmatrix}
\]

\[-2* \text{loglikelihood(IGLS Deviance)} = 2225.881 (891 of 920 cases in use)\]

\[
\text{ICC} = \frac{.346}{.346 + .646} = .348
\]
History

\[ Z_{ctvhis}^{2}_{ijk} \sim N(\lambda B, \Omega) \]

\[ Z_{ctvhis}^{2}_{ijk} = \beta_{0ijk} \text{CONS} \]

\[ \beta_{0ijk} = 0.008(0.106) + \nu_{0k} + \mu_{0jk} + e_{0ijk} \]

\[
\begin{bmatrix}
\nu_{0k}
\end{bmatrix} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix}
0.000(0.000)
\end{bmatrix}
\]

\[
\begin{bmatrix}
\mu_{0jk}
\end{bmatrix} \sim N(0, \Omega_{\mu}) : \Omega_{\mu} = \begin{bmatrix}
0.346(0.091)
\end{bmatrix}
\]

\[
\begin{bmatrix}
e_{0ijk}
\end{bmatrix} \sim N(0, \Omega_{e}) : \Omega_{e} = \begin{bmatrix}
0.646(0.031)
\end{bmatrix}
\]

\[-2*\text{log likelihood}(IGLS Deviance) = 2283.054(913 of 920 cases in use)\]

\[ \text{ICC} = \frac{0.346}{0.346 + 0.646} = 0.348 \]
4. Calculating ICC for Blanton et al. (1999) for unmodified grades

(Study 2a)

Geography

\[ \text{zaar}_r^2_{ijk} \sim N(XB, \Omega) \]

\[ \text{zaar}_r^2_{ijk} = \beta_{0ijk}\text{CONS} \]

\[ \beta_{0ijk} = 0.000(0.032) + \nu_{0k} + \nu_{0jk} + \psi_{0ijk} \]

\[ \begin{bmatrix} \nu_{0k} \\ \nu_{0jk} \end{bmatrix} \sim N(0, \Omega) : \Omega = \begin{bmatrix} 0.000(0.000) \\ 0.000(0.000) \end{bmatrix} \]

\[ \begin{bmatrix} \psi_{0ijk} \end{bmatrix} \sim N(0, \Omega_{\psi}) : \Omega_{\psi} = \begin{bmatrix} 0.964(0.045) \end{bmatrix} \]

\[-2*\log\text{likelihood(TGLS Deviance)} = 2557.372(913 \text{ of } 920 \text{ cases in use})\]

Biology

\[ \text{zbio}_r^2_{ijk} \sim N(XB, \Omega) \]

\[ \text{zbio}_r^2_{ijk} = \beta_{0ijk}\text{CONS} \]

\[ \beta_{0ijk} = 0.000(0.033) + \nu_{0k} + \nu_{0jk} + \psi_{0ijk} \]

\[ \begin{bmatrix} \nu_{0k} \\ \nu_{0jk} \end{bmatrix} \sim N(0, \Omega) : \Omega = \begin{bmatrix} 0.000(0.000) \\ 0.000(0.000) \end{bmatrix} \]

\[ \begin{bmatrix} \psi_{0ijk} \end{bmatrix} \sim N(0, \Omega_{\psi}) : \Omega_{\psi} = \begin{bmatrix} 0.964(0.045) \end{bmatrix} \]

\[-2*\log\text{likelihood(TGLS Deviance)} = 2554.532(912 \text{ of } 920 \text{ cases in use})\]
\[ z_{\text{eng}, r2_{ijk}} \sim N(\mathbf{X}_i \mathbf{B}_i, \Omega) \]
\[ z_{\text{eng}, r2_{ijk}} = \beta_{0ijk} \text{CONS} \]
\[ \beta_{0ijk} = 0.000(0.033) + \nu_{0k} + \mu_{0jk} + \epsilon_{0ijk} \]

\[
\begin{bmatrix}
\nu_{0k} \\
\mu_{0jk} \\
\epsilon_{0ijk}
\end{bmatrix} \sim N(0, \Omega_v) \quad : \quad \Omega_v = \begin{bmatrix}
0.000(0.000)
\end{bmatrix}
\]

\[
\begin{bmatrix}
\nu_{ijk}
\end{bmatrix} \sim N(0, \Omega_u) \quad : \quad \Omega_u = \begin{bmatrix}
0.000(0.000)
\end{bmatrix}
\]

\[
\begin{bmatrix}
\epsilon_{0ijk}
\end{bmatrix} \sim N(0, \Omega_\varepsilon) \quad : \quad \Omega_\varepsilon = \begin{bmatrix}
0.964(0.045)
\end{bmatrix}
\]

\[-2*\text{loglikelihood(IQLS Deviance)} = 2551.694(911 \text{ of } 920 \text{ cases in use})\]

---

**French**

\[ z_{\text{fra}, r2_{ijk}} \sim N(\mathbf{X}_i \mathbf{B}_i, \Omega) \]
\[ z_{\text{fra}, r2_{ijk}} = \beta_{0ijk} \text{CONS} \]
\[ \beta_{0ijk} = 0.000(0.033) + \nu_{0k} + \mu_{0jk} + \epsilon_{0ijk} \]

\[
\begin{bmatrix}
\nu_{0k} \\
\mu_{0jk} \\
\epsilon_{0ijk}
\end{bmatrix} \sim N(0, \Omega_v) \quad : \quad \Omega_v = \begin{bmatrix}
0.000(0.000)
\end{bmatrix}
\]

\[
\begin{bmatrix}
\nu_{ijk}
\end{bmatrix} \sim N(0, \Omega_u) \quad : \quad \Omega_u = \begin{bmatrix}
0.000(0.000)
\end{bmatrix}
\]

\[
\begin{bmatrix}
\epsilon_{0ijk}
\end{bmatrix} \sim N(0, \Omega_\varepsilon) \quad : \quad \Omega_\varepsilon = \begin{bmatrix}
0.964(0.045)
\end{bmatrix}
\]

\[-2*\text{loglikelihood(IQLS Deviance)} = 2548.855(910 \text{ of } 920 \text{ cases in use})\]
History

$$zges_{-r2_{ijk}} \sim N(XB, \Omega)$$

$$zges_{-r2_{ijk}} = \beta_{0ijk} \text{CONS}$$

$$\beta_{0ijk} = 0.000(0.032) + \nu_{0k} + \nu_{0jk} + \nu_{0ijk}$$

$$\begin{bmatrix} \nu_{0k} \\ \nu_{0jk} \end{bmatrix} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.000(0.000) \\ 0.000(0.000) \end{bmatrix}$$

$$\begin{bmatrix} \nu_{0ijk} \end{bmatrix} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.964(0.045) \end{bmatrix}$$

$$-2*\text{loglikelihood (IGLS Deviance)} = 2557.370 (913 \text{ of 920 cases in use})$$

Dutch

$$zned_{-r2_{ijk}} \sim N(XB, \Omega)$$

$$zned_{-r2_{ijk}} = \beta_{0ijk} \text{CONS}$$

$$\beta_{0ijk} = 0.000(0.032) + \nu_{0k} + \nu_{0jk} + \nu_{0ijk}$$

$$\begin{bmatrix} \nu_{0k} \\ \nu_{0jk} \end{bmatrix} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.000(0.000) \\ 0.000(0.000) \end{bmatrix}$$

$$\begin{bmatrix} \nu_{0ijk} \end{bmatrix} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.964(0.045) \end{bmatrix}$$

$$-2*\text{loglikelihood (IGLS Deviance)} = 2557.371 (913 \text{ of 920 cases in use})$$
Math

\( zwis_{r2ijk} \sim N(\Theta, \Omega) \)

\( zwis_{r2ijk} = \beta_{0ij} \text{CONS} \)

\[ \beta_{0ijk} = 0.000(0.033) + \nu_{0k} + \nu_{0jk} + \varepsilon_{0ijk} \]

\[
\begin{bmatrix}
\nu_{0k}
\end{bmatrix} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.000(0.000) \end{bmatrix}
\]

\[
\begin{bmatrix}
\nu_{0jk}
\end{bmatrix} \sim N(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.000(0.000) \end{bmatrix}
\]

\[
\begin{bmatrix}
\varepsilon_{0ijk}
\end{bmatrix} \sim N(0, \Omega_{\varepsilon}) : \Omega_{\varepsilon} = \begin{bmatrix} 0.964(0.046) \end{bmatrix}
\]

-2*loglikelihood(IQLS Deviance) = 2495.960 (891 of 920 cases in use)

Note. All class and school random effects = zero, so no calculations necessary. ICC = 0.
5. Calculating ICC for Huguet et al. expanded data (Study 2b)

**French**

\[ zt2 \mu_{ijk} \sim \mathcal{N}(XB, \Omega) \]

\[ zt2 \mu_{ijk} = \beta_{\mu ij k} \text{CONS} \]

\[ \beta_{\mu ij k} = -0.010(0.067) + \nu_{0k} + \mu_{0jk} + \varepsilon_{0ijk} \]

\[
\begin{bmatrix}
\nu_{0k}
\end{bmatrix} \sim \mathcal{N}(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.010(0.027) \end{bmatrix}
\]

\[
\begin{bmatrix}
\mu_{0jk}
\end{bmatrix} \sim \mathcal{N}(0, \Omega_{\mu}) : \Omega_{\mu} = \begin{bmatrix} 0.153(0.045) \end{bmatrix}
\]

\[
\begin{bmatrix}
\varepsilon_{0ijk}
\end{bmatrix} \sim \mathcal{N}(0, \Omega_{\epsilon}) : \Omega_{\epsilon} = \begin{bmatrix} 0.835(0.036) \end{bmatrix}
\]

\[-2\times \text{loglikelihood} (IGLS Deviance) = 3139.230 (1149 of 1156 cases in use)\]

School ICC = \[
\frac{0.010}{0.010 + 0.153 + 0.835} = 0.010
\]

Class ICC = \[
\frac{0.153}{0.010 + 0.153 + 0.835} = 0.153
\]

**Maths**

\[ zt2 \mu_{ijk} \sim \mathcal{N}(XB, \Omega) \]

\[ zt2 \mu_{ijk} = \beta_{\mu ij k} \text{CONS} \]

\[ \beta_{\mu ij k} = -0.105(0.079) + \nu_{0k} + \mu_{0jk} + \varepsilon_{0ijk} \]

\[
\begin{bmatrix}
\nu_{0k}
\end{bmatrix} \sim \mathcal{N}(0, \Omega_{\nu}) : \Omega_{\nu} = \begin{bmatrix} 0.087(0.039) \end{bmatrix}
\]

\[
\begin{bmatrix}
\mu_{0jk}
\end{bmatrix} \sim \mathcal{N}(0, \Omega_{\mu}) : \Omega_{\mu} = \begin{bmatrix} 0.028(0.016) \end{bmatrix}
\]

\[
\begin{bmatrix}
\varepsilon_{0ijk}
\end{bmatrix} \sim \mathcal{N}(0, \Omega_{\epsilon}) : \Omega_{\epsilon} = \begin{bmatrix} 0.843(0.036) \end{bmatrix}
\]

\[-2\times \text{loglikelihood} (IGLS Deviance) = 3133.147 (1154 of 1156 cases in use)\]

School ICC = \[
\frac{0.087}{0.087 + 0.028 + 0.843} = 0.0908
\]

Class ICC = \[
\frac{0.028}{0.087 + 0.028 + 0.843} = 0.0292
\]
History

$z_{2h_{ijk}} \sim N(\beta, \Omega)$

$z_{2h_{ijk}} = \beta_{ijk} + \nu_{ijk} + \mu_{ijk} + \epsilon_{ijk}$

$\beta_{ijk} = -0.094(0.075) + \nu_{ijk} + \mu_{ijk} + \epsilon_{ijk}$

$[\nu_{ijk}] \sim N(0, \Omega_\nu) : \Omega_\nu = \begin{bmatrix} 0.067 & 0.035 \end{bmatrix}$

$[\mu_{ijk}] \sim N(0, \Omega_\mu) : \Omega_\mu = \begin{bmatrix} 0.043 & 0.020 \end{bmatrix}$

$[\epsilon_{ijk}] \sim N(0, \Omega_\epsilon) : \Omega_\epsilon = \begin{bmatrix} 0.852 & 0.036 \end{bmatrix}$

$-2 \times \text{loglikelihood (IGLS Deviance)} = 3132.493 (1148 \text{ of 1156 cases in use})$

School ICC = $\frac{.067}{.067 + .043 + .852} = .0696$

Class ICC = $\frac{.028}{.067 + .043 + .852} = .04469$
This appendix provides details of analyses performed to assess that multiple imputation was appropriate in Study 3. The questionnaire that formed the basis of Study 3 is also included.

**Appropriateness of Multiple Imputation to Cater for Missing Data**

To assess the appropriateness of multiple imputation to cater for missing data, similar analyses to those performed in Study 3 (Chapter 6) were conducted on data where the comparison choice information was not imputed. These analyses, for both the objective and subjective comparison measures, revealed a similar pattern of results to those reported in Chapter 6 (see Tables C.1 and C.2). Therefore, it was concluded that imputing the missing data was an appropriate strategy to adopt.

Table C.1. *Objective Comparison Choice as a Moderator of the BFLPE when Comparison Choice Test Scores Not Imputed*

<table>
<thead>
<tr>
<th></th>
<th>French Self-concept</th>
<th>Math Self-concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind Ability</td>
<td>0.495*</td>
<td>0.657*</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Class-Average</td>
<td>-0.505*</td>
<td>-0.613*</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>Comparison</td>
<td>0.102*</td>
<td>0.113*</td>
</tr>
<tr>
<td>Other</td>
<td>(0.031)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Comp X Class</td>
<td>0.133*</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.034</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>0.024</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Class</td>
<td>0.054*</td>
<td>0.035*</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Individual</td>
<td>0.774*</td>
<td>0.665*</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.024)</td>
</tr>
</tbody>
</table>

*Note.* Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). All parameter estimates are standardised coefficients. SE in parentheses. Comp X Class = Comparison other Ability X Class-average Ability.
Table C.2. *Subjective Comparative Choice as a Moderator of the BFLPE when Comparative Choice Scores Not Imputed (“The Same” is the reference category)*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>French Self-concept</th>
<th>Math Self-concept</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual Ability</strong></td>
<td>0.415* (0.026)</td>
<td>0.604* (0.025)</td>
</tr>
<tr>
<td><strong>Class-Average Ability</strong></td>
<td>-0.418* (0.073)</td>
<td>-0.497* (0.071)</td>
</tr>
<tr>
<td><strong>No Choice</strong></td>
<td>-0.265* (0.060)</td>
<td>-0.243* (0.060)</td>
</tr>
<tr>
<td><strong>Much Worse</strong></td>
<td>-0.857* (0.080)</td>
<td>-0.734* (0.092)</td>
</tr>
<tr>
<td><strong>Slightly Worse</strong></td>
<td>-0.343* (0.054)</td>
<td>-0.228* (0.050)</td>
</tr>
<tr>
<td><strong>Slightly Better</strong></td>
<td>0.188* (0.055)</td>
<td>0.184* (0.049)</td>
</tr>
<tr>
<td><strong>Much Better</strong></td>
<td>0.386* (0.082)</td>
<td>0.500* (0.072)</td>
</tr>
<tr>
<td><strong>NC X Class</strong></td>
<td>-0.109 (0.089)</td>
<td>-0.113 (0.094)</td>
</tr>
<tr>
<td><strong>MW X Class</strong></td>
<td>0.034 (0.121)</td>
<td>-0.009 (0.124)</td>
</tr>
<tr>
<td><strong>SW X Class</strong></td>
<td>0.061 (0.083)</td>
<td>0.066 (0.077)</td>
</tr>
<tr>
<td><strong>SB X Class</strong></td>
<td>-0.003 (0.084)</td>
<td>0.113 (0.076)</td>
</tr>
<tr>
<td><strong>MB X Class</strong></td>
<td>0.018 (0.112)</td>
<td>0.086 (0.101)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.106* (0.049)</td>
<td>0.033 (0.051)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>School</th>
<th>Class</th>
<th>Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.011</td>
<td>0.063* (0.15)</td>
<td>0.693* (0.022)</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>

**Note:** Parameter estimates are significant at 0.05 level (denoted by *) when they differ from zero by more than two standard errors (SEs). All parameter estimates are standardised coefficients. SE in brackets. NC X Class = No choice X Class-average Ability; MW X Class = Much worse X Class-average Ability; SW X Class = Slightly worse X Class-average Ability; SB X Class = Slightly better X Class-average Ability; MB X Class = Much better X Class-average Ability.
**QUESTIONAIRRE - STUDY 3**

<table>
<thead>
<tr>
<th><strong>COMPARISON-LEVEL CHOICE</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. With which student (boy or girl) do you prefer to compare your marks in math? Write his/her identity number: . . .</td>
</tr>
<tr>
<td>I do not know □</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>IDENTIFICATION TO THE TARGET (A)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Has this student the same marks as yours in math?</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>never</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>COMPARATIVE EVALUATION (Oneyeself)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>3. How much better/worse are you in math compared to most of your classmates?</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>much worse</td>
</tr>
</tbody>
</table>

---

**Do you agree with the following sentences?**

<table>
<thead>
<tr>
<th><strong>PERCEPTION OF ACADEMIC CONTROL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>4. You can make progress in math if you make efforts</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>strongly disagree</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>IMPLICIT THEORIES OF INTELLIGENCE SCALE FOR CHILDREN –SELF FORM</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>5. No matter who you are, you can change your ability in math a lot (incremental item)</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>strongly disagree</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Your ability in math is something about you that you can’t change very much (entity item)</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>strongly disagree</td>
</tr>
</tbody>
</table>
7. You can always greatly change your ability in Math a lot (*incremental item*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

8. You can learn new things, but you can’t really change your basic ability in math (*entity item*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

9. No matter how much ability in math you have, you can always change it quite a bit (*incremental item*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

10. You have a certain ability in math, and you really can’t do much to change it (*entity item*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

**SIX-ITEM SELF-CONCEPT SCALE (SPECIFIC SCHOOL DOMAINS)**

11. You have always done well in math

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

12. Compared to others your age you are good at math

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

13. Work in math classes is easy for you

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

14. You are hopeless when it comes to math (*reverse scored*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>
15. You get good marks in math

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

16. You learn things quickly in math

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

**ACHIEVEMENT GOAL QUESTIONNAIRE**

17. In math, it is important for you to understand what the teacher explains in class as thoroughly as possible *(mastery approach)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

18. In math, your goal in class is to get a better grade than most of the other students *(performance approach)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

19. In math, you worry that you may not learn all that you possibly can in class *(mastery avoidance)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

20. In math, your goal in class is to avoid performing poorly *(performance avoidance)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

21. In math, it is important for you to do better than other students *(performance approach)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>agree</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>
### 22. In math, you are often concerned that you may not learn all there is to learn in class (*mastery avoidance*)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
<td></td>
</tr>
</tbody>
</table>

### 23. In math, you desire to completely master the material presented in class (*mastery approach*)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
<td></td>
</tr>
</tbody>
</table>

### 24. In math, your fear of performing poorly in class is often what motivates you (*performance avoidance*)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
<td></td>
</tr>
</tbody>
</table>

### 25. In math, you want to learn as much as possible in class (*mastery approach*)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
<td></td>
</tr>
</tbody>
</table>

### 26. In math, sometimes you are afraid that you may not understand what the teacher explains in class as thoroughly as possible (*mastery avoidance*)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
<td></td>
</tr>
</tbody>
</table>

### 27. In math, it is important for you to do well compared to others in your class (*performance approach*)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
<td></td>
</tr>
</tbody>
</table>

### 28. In math, you just want to avoid doing poorly in class (*performance avoidance*)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
<td></td>
</tr>
</tbody>
</table>
### COMPARISON-LEVEL CHOICE

**29.** With which student (boy or girl) do you prefer to compare your marks in French? Write his/her identity number: ..

I do not know [ ]

### IDENTIFICATION TO THE TARGET (A)

**30.** Has this student the same marks as yours in French? :

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>never</td>
<td>Sometimes</td>
<td>one time out of two</td>
<td>often</td>
<td>always</td>
</tr>
</tbody>
</table>

### COMPARATIVE EVALUATION (ONESELF)

**31.** How much better/worse are you in French compared to most of your classmates?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>much worse</td>
<td>slightly worse</td>
<td>the same</td>
<td>slightly better</td>
<td>much better</td>
</tr>
</tbody>
</table>

### Do you agree with the following sentences?

### PERCEPTION OF ACADEMIC CONTROL

**32.** You can make progress in French if you make efforts

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

### IMPLICIT THEORIES OF INTELLIGENCE SCALE FOR CHILDREN – SELF FORM

**33.** No matter who you are, you can change your ability in French a lot *(incremental item)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

**34.** Your ability in French is something about you that you can’t change very much *(entity item)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>
### 35. You can always greatly change your ability in French a lot (incremental item)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

### 36. You can learn new things, but you can’t really change your basic ability in French (entity item)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

### 37. No matter how much ability in French you have, you can always change it quite a bit (incremental item)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

### 38. You have a certain ability in French, and you really can’t do much to change it (entity item)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

---

**SIX-ITEM SELF-CONCEPT SCALE (SPECIFIC SCHOOL DOMAINS)**

### 39. You have always done well in French

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

### 40. Compared to others your age you are good at French

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

### 41. Work in French classes is easy for you

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

### 42. You are hopeless when it comes to French (reverse scored)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>
43. You get good marks in French

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

44. You learn things quickly in French

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

**ACHIEVEMENT GOAL QUESTIONNAIRE**

45. In French, it is important for you to understand what the teacher explains in class as thoroughly as possible (*mastery approach*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

46. In French, your goal in class is to get a better grade than most of the other students (*performance approach*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

47. In French, you worry that you may not learn all that you possibly can in class (*mastery avoidance*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

48. In French, your goal in class is to avoid performing poorly (*performance avoidance*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

49. In French, it is important for you to do better than other students (*performance approach*)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>
50. In French, you are often concerned that you may not learn all there is to learn in class *(mastery avoidance)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>2</td>
<td>3</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

51. In French, you desire to completely master the material presented in class *(mastery approach)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>2</td>
<td>3</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

52. In French, your fear of performing poorly in class is often what motivates you *(performance avoidance)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>2</td>
<td>3</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

53. In French, you want to learn as much as possible in class *(mastery approach)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>2</td>
<td>3</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

54. In French, sometimes you are afraid that you may not understand what the teacher explains in class as thoroughly as possible *(mastery avoidance)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>2</td>
<td>3</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

55. In French, it is important for you to do well compared to others in your class *(performance approach)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>2</td>
<td>3</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

56. In French, you just want to avoid doing poorly in class *(performance avoidance)*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly disagree</td>
<td>2</td>
<td>3</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>
### Class Status Scale

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>57. The other students and teachers of the school think that your class is one of the best academically</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>58. Academically you are proud to be part of this class</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>59. You find it academically rewarding to be part of this class</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>60. The academic standard of this class is disappointing to you</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

### Generalised vs. Specific Others

What do you do when you receive a mark on a test?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>61. You compare you mark with someone who perform generally as yourself in the same domain (specific other)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>62. You compare your mark with someone who perform generally better than yourself (specific other)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>63. You compare your mark with someone who perform generally worse than yourself (specific other)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>64. You compare your mark with that of a friend in the class (specific other)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>65. You look how many students performed better than you (generalised other)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>66. You look how many students performed lower than you (generalised other)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>67. You look how many students obtained the same mark as yours (generalised other)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>68. You want to know which is the worst mark on this test (extreme range)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>69. You want to know which is the best mark on this test (extreme range)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>70. You compare your mark with the class average (generalised other)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>71. You pay attention to what the teacher thinks about your mark (objective standard)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td>72. You look if you reached 10 (the acceptable mark) (objective standard)</td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>
73. You compare your mark with the last one you obtained in the same domain (temporal comparison)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

**School Status Scale**

74. Students and teachers from other schools think that your school is one of the best academically

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

75. Academically you are proud to be part of this school

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

76. You find it academically rewarding to be part of this school

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

77. The academic standard of this school is disappointing to you

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strongly disagree</td>
<td>disagree</td>
<td>mostly disagree</td>
<td>mostly agree</td>
<td>agree</td>
<td>strongly agree</td>
</tr>
</tbody>
</table>

**Explicit Comparative Evaluation**

78. On page 1, you wrote the identity number of a student with whom you compare your marks in math:

Compared to him (her), in math, you are:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>much worse</td>
<td>slightly worse</td>
<td>the same</td>
<td>slightly better</td>
<td>much better</td>
</tr>
</tbody>
</table>

If you did not choose anyone, tick off this square:

79. On page 1, you wrote the identity number of a student with whom you compare your marks in French:

Compared to him (her), in French, you are:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>much worse</td>
<td>slightly worse</td>
<td>the same</td>
<td>slightly better</td>
<td>much better</td>
</tr>
</tbody>
</table>

If you did not choose anyone, tick off this square:
THANKS YOU FOR YOUR HELP!
REFERENCES


335


Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (in review). Social comparison and big-fish-little-pond effects on self-concept and efficacy perceptions: Role of generalised and specific others.


Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society, 54, 426-482.


348

