Degrees of autonomy in foreign language learning

by

Carol J. Everhard-Theophilidou

A thesis submitted for the degree of

Doctor of Philosophy

to the

Department of Theoretical and Applied Linguistics

School of English

Aristotle University of Thessaloniki

VOLUME 1

January 2012
To my late parents, Ada and William, who made me who I am, to my late parents-in-law, Chryssoula and Theophilos, who helped me integrate in Greek society and encouraged me in pursuing my professional and academic goals, to my late sister Patricia, who left us much too early, to my brother Bill, who has always been there for me, and last, but certainly not least, to my husband George and my daughter Christina, for demanding the best of me and being the bountiful treasures that they are.
ABSTRACT

A five-year study related to autonomy in foreign language learning, which in this instance was English as a Foreign Language, was implemented in the School of English (SOE) of the Faculty of Philosophy, Aristotle University of Thessaloniki and involved the investigation of peer-assessment and self-assessment of writing and speaking skills as a means to promote greater autonomy in language learning. The study, entitled the Assessment for Autonomy Research Project (AARP), consisted of three phases extending over the five academic years 2005–2010: a Preliminary Study (2005-2006), the Main Study (2006-2009) and a Post-Study (2009-2010). The three study phases were conducted with students on the SOE 1st Year course, Language Mastery I (LM I). Subjects used in the AARP study were convenience samples from the Instructor-Researcher’s own LM I course groups.

Quantitative data concerning peer-, self- and instructor assessment of both writing skills and speaking skills was gathered by means of criterial assessment check-lists for two home writing assignments and one in-class oral presentation, prepared at home. The same criteria were used by all participants, facilitating triangulation of results and their analysis, while qualitative data was gathered by means of assessment questionnaires, with responses given on a Likert scale, which could be analysed quantitatively, as well as offering comments which could be selected and presented thematically.

The research is based on the working hypothesis that by adopting a model for language learning which allows for variation and modification in language learner behaviour, on a continuum between heteronomy and autonomy, and by accepting a definition of autonomy which allows, at the same time, for differences in disposition and uptake, and for both universal principles and local conditions, it is possible to view engagement with autonomy in language learning as varying or fluctuating in degrees.

The research questions aimed to establish if learners could assess the oral and writing skills of their peers and of themselves with objectivity and reliability, using the same predetermined criteria, and to see if there was evidence that learners would assume ownership of the assessment criteria checklists and use them to exercise judgement in an atmosphere of cooperation and trust.
Although the period of student involvement in peer- and self-assessment processes was relatively short, the results seem to indicate that it was possible for students to overcome their existing preconceptions and beliefs about learning and engage in reflection and the necessary critical and criterial thinking demanded in the production and assessment of assignments. Thus, they accepted a shifting and recalibration of power over assessment and moved towards assessment literacy and away from the heteronomy to which they were accustomed. Using peer-assessment as a stepping-stone towards self-assessment, learners, for the most part were able to assume greater responsibility and self-direction and thus exercise a greater degree of autonomy, developing skills which will be useful to them throughout a lifetime of living, learning, and, in the majority of cases, language teaching.
ACKNOWLEDGEMENTS

This thesis is dedicated to those who offered support and encouragement during the many trials involved in completing the research and writing up of my findings. The sacrifices in this process, unfortunately, were not exclusively mine and apologies are due to friends and family who did not always have my company, my moral support or my full attention when they perhaps needed it most.

I am indebted to the Faculty of the School of English for their patience, to Hon. Prof. Stathis Efstatiadis for initially accepting me as a PhD student, and to former Associate Professor Niovi Antonopoulou who took over as my Main Supervisor, but was consequently obliged to stand down. My indebtedness to Prof. Antonopoulou is indeed immeasurable, since, at her suggestion, the implementation of self-assessment and the use of assessment criteria check-lists in my teaching practices, helped me uncover an essential missing link in the autonomy conundrum.

Assistant Professor Areti-Maria Sougari, in spite of her already heavy workload and the difficult circumstances, saved the day, by undertaking in the late stages to become my Main Supervisor, and helped me, through her critical wisdom, her wealth of knowledge and experience, as well as her exacting standards, see this very demanding task through to completion. Words cannot express my gratitude to her.

The two key stalwarts in the progress of this research have been Professor Angeliki Athanassiadou and Associate Professor Michalis Milapides, who remained, from beginning to end, unwavering in their support. Without their meticulous and conscientious guidance, this research would never have been completed.

I also acknowledge my indebtedness to Dr Dimitris Petridis for his patience and guidance in the analysis of my statistical data and to Dr George Triantafyllakos for his initial help with figure and table templates.

Likewise, my gratitude to the many students of the School of English who passed through my hands is inestimable. Special thanks go to the participants in the research project. Their daring and willingness to explore unknown terrain, their perseverance, and their insightful views on the research subject-matter never cease to amaze me.

While much work remains to be done on the subject of autonomy in language learning and on the link between assessment and autonomy, it is hoped that some of the findings and insights provided here will be of value to present and future researchers in the field.
**CONVENTIONS USED IN THE THESIS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AARP</td>
<td>Assessment for Autonomy Research Project (2005-2010)</td>
</tr>
<tr>
<td>ALL</td>
<td>Autonomous Language Learning</td>
</tr>
<tr>
<td>ALTE</td>
<td>Association of Language Testers in Europe</td>
</tr>
<tr>
<td>ANOVA</td>
<td>One-way Analysis of Variance</td>
</tr>
<tr>
<td>AUTh</td>
<td>Aristotle University of Thessaloniki</td>
</tr>
<tr>
<td>CEFR</td>
<td>Common European Framework of Reference</td>
</tr>
<tr>
<td>CG</td>
<td>Control Group</td>
</tr>
<tr>
<td>CLL</td>
<td>Communicative Language Learning</td>
</tr>
<tr>
<td>CLT</td>
<td>Communicative Language Teaching</td>
</tr>
<tr>
<td>COE</td>
<td>Council of Europe</td>
</tr>
<tr>
<td>CRAPEL</td>
<td>Centre de Recherches et d’Applications en Langues, Nancy, France</td>
</tr>
<tr>
<td>EAP</td>
<td>English for Academic Purposes</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a Foreign Language</td>
</tr>
<tr>
<td>EG</td>
<td>Experimental Group</td>
</tr>
<tr>
<td>ELP</td>
<td>English Language Portfolio</td>
</tr>
<tr>
<td>ELT</td>
<td>English Language Teaching</td>
</tr>
<tr>
<td>FL</td>
<td>Foreign Language</td>
</tr>
<tr>
<td>FLL</td>
<td>Foreign Language Learning</td>
</tr>
<tr>
<td>FLT</td>
<td>Foreign Language Teaching</td>
</tr>
<tr>
<td>HE</td>
<td>Higher Education</td>
</tr>
<tr>
<td>I-A</td>
<td>Instructor Assessment</td>
</tr>
<tr>
<td>IE</td>
<td>Intervention Exercise</td>
</tr>
<tr>
<td>I:P</td>
<td>Ratio of Instructor-assessment as compared with Peer-assessment</td>
</tr>
<tr>
<td>I:R</td>
<td>Instructor-Researcher</td>
</tr>
<tr>
<td>I:S</td>
<td>Ratio of Instructor-assessment as compared with Self-assessment</td>
</tr>
<tr>
<td>L2</td>
<td>Second Language</td>
</tr>
<tr>
<td>LM I</td>
<td>Language Mastery I course, taught in the 1st semester</td>
</tr>
<tr>
<td>MFL</td>
<td>Modern Foreign Languages</td>
</tr>
<tr>
<td>n.s.</td>
<td>Non-significant differences in One-way ANOVA</td>
</tr>
<tr>
<td>OE</td>
<td>Oral Examiner</td>
</tr>
<tr>
<td>OPT</td>
<td>Oxford Placement Test</td>
</tr>
<tr>
<td>P-A</td>
<td>Peer-assessment</td>
</tr>
<tr>
<td>P-I</td>
<td>Peer-Instructor Pearson correlation coefficient</td>
</tr>
<tr>
<td>P-S</td>
<td>Peer-Self Pearson correlation coefficient</td>
</tr>
<tr>
<td>P:S</td>
<td>Ratio of Peer-assessment as compared with Self-assessment</td>
</tr>
<tr>
<td>RC</td>
<td>Resource Centre</td>
</tr>
<tr>
<td>S-A</td>
<td>Self-assessment</td>
</tr>
<tr>
<td>SALL</td>
<td>Self-access Language Learning</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>S.E.M.</td>
<td>Standard Error of Measurement</td>
</tr>
<tr>
<td>S-I</td>
<td>Self-Instructor Pearson correlation coefficient</td>
</tr>
<tr>
<td>sig.</td>
<td>Significant differences in One-way ANOVA</td>
</tr>
<tr>
<td>SOE</td>
<td>School of English, Aristotle University of Thessaloniki</td>
</tr>
<tr>
<td>S-P</td>
<td>Self-Peer</td>
</tr>
<tr>
<td>TEA</td>
<td>Testing, Evaluation and Assessment</td>
</tr>
<tr>
<td>TL</td>
<td>Team Leader</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS

ABSTRACT i
ACKNOWLEDGEMENTS iii
CONVENTIONS USED IN THE THESIS iv
TABLE OF CONTENTS v
LIST OF FIGURES ix
LIST OF TABLES x

CHAPTER 1 INTRODUCTION TO THE THESIS
1.0 Introduction 1
1.1 The backdrop to the AARP 1
1.2 The significance of the AARP 3
1.3 Preliminary research questions 5
1.4 The aims and scope of the research 7
1.5 Organisation of the thesis 9
1.6 Concluding remarks 10

CHAPTER 2 AUTONOMY IN LANGUAGE LEARNING: AN OVERVIEW
2.0 Introduction 11
2.1 Why autonomy? 12
  2.1.1 The implications of promoting autonomy 15
2.2 Autonomy vs. heteronomy 18
2.3 Some key elements in the promotion of autonomy 23
  2.3.1 Identity 26
  2.3.2 Reflection 29
  2.3.3 Ownership 31
  2.3.4 Self-determination 35
2.4 Areas of consensus on autonomy 39
2.5 Defining autonomy 44
  2.5.1 Multiple interpretations and definitions 45
  2.5.2 The AARP definition of autonomy 51
  2.5.3 The answer to the 1st preliminary research question 51
2.6 Finding a model for the AARP 52
  2.6.1 Models of autonomy 53
  2.6.2 The proposed AARP model 58
  2.6.3 The answer to the 2nd preliminary research question 63
2.7 Concluding remarks 63

CHAPTER 3 ASSESSMENT FOR AUTONOMY: AN OVERVIEW
3.0 Introduction 64
3.1 Testing, evaluation and assessment (TEA) 64
  3.1.1 TEA and its influence on learning 65
  3.1.2 TEA and its influence on language learning 67
  3.1.3 Who controls TEA? 70
  3.1.4 TEA and its influence on learner beliefs and attitudes 71
3.1.5 TEA and its influence on learner autonomy
3.1.6 The need for assessment literacy

3.2 Alternative assessment, the COE and the CEFR
3.2.1 The contribution of Holec to the field of assessment
3.2.2 The foresight of Oscarson (formerly Oskarsson)
3.2.3 Dickinson’s contribution to the field
3.2.4 Little’s Council of Europe work
3.2.5 Kohonen’s experiential approach

3.3 Two world-leading proponents of self- and peer-assessment
3.3.1 Boud, autonomy and self-assessment
3.3.2 Falchikov and apprenticeship learning

3.4 An overview of research studies most relevant to the AARP
3.4.1 Studies of writing skills assessment in HE
3.4.2 Studies of oral skills assessment in HE
3.4.3 Dual assessment of oral and writing skills in HE

3.5 Concluding remarks

CHAPTER 4 AARP RESEARCH METHODOLOGY
4.0 Introduction
4.1 The AARP setting
4.2 Research hypothesis and research questions
4.3 Research design
4.3.1 The research methods used
4.3.2 The research instruments used
4.3.2.1 Criteria checklists
4.3.2.2 Assessment questionnaires
4.4 Background to the AARP
4.4.1 The academic calendar and the AARP
4.4.2 The assigning of students to groups
4.4.3 Factors beyond the I-R’s control
4.4.4 Gathering data about the participants
4.4.5 Profile cards and contracts
4.5 Assessment procedures on Language Mastery I
4.5.1 The assessment criteria
4.5.2 The assessment cycle
4.6 Pedagogical procedures for writing and oral production
4.6.1 Procedures for the 1st Home writing assignment
4.6.1.1 AARP Post-Study 1st Home writing assignment Intervention exercise
4.6.2 Procedures for the 2nd Home writing assignment
4.6.2.1 AARP Main Study 2nd Home writing assignment variation (2006-2007)
4.6.2.2 AARP Post-Study 2nd Home writing assignment non-intervention
4.6.3 Procedures for the Oral presentations and assessment
4.6.3.1 AARP Post-Study oral assignment intervention
4.7 Analysis of the assessment data
4.7.1 Analysis of data gathered from use of criterial checklists
4.7.2 Analysis of data gathered from use of assessment questionnaires
4.8 Concluding remarks
CHAPTER 5  AARP PRELIMINARY (PRE-), MAIN AND POST-STUDY RESULTS

5.0       Introduction  165
5.1       The AARP Preliminary Study (Pre-Study)  166
  5.1.1   Breakdown of participants in the AARP Pre-Study, 2005-2006  166
  5.1.2   AARP Pre-Study 1st Home Writing Assignment assessment results  169
  5.1.3   AARP Pre-Study 2nd Home Writing Assignment assessment results  171
  5.1.4   AARP Pre-Study Oral Assignment assessment results  172
  5.1.5   AARP Pre-Study Questionnaire results  174
  5.1.6   AARP Pre-Study - Overview of outcome  176
5.2       The AARP Main Study Results  177
  5.2.1   Breakdown of participants in the AARP Main Study, 2006-2009  178
  5.2.2   AARP Main Study 1st Home Writing Assignment assessment results  181
  5.2.3   AARP Main Study 2nd Home Writing Assignment assessment results  184
    5.2.3.1   AARP Main Study (2006-2007) Variation in Self-assessment of the 2nd Writing Assignment  187
  5.2.4   AARP Main Study Oral Assignment assessment results  192
  5.2.5   AARP Main Study Questionnaire results  196
  5.2.6   AARP Main Study - Overview of outcome  199
5.3       The AARP Post-Study Results  201
  5.3.1   Breakdown of participants in the AARP Post-Study, 2009-2010  202
  5.3.2   AARP Post-Study Assessment results  204
    5.3.2.1   AARP Post-Study 1st Writing Assignment Intervention Exercise results  205
    5.3.2.2   AARP Post-Study (Actual) 1st Writing Assignment assessment results  206
  5.3.3   AARP Post-Study 2nd Home Writing Assignment assessment results  209
  5.3.4   AARP Post-Study Oral Assignment assessment and Intervention
    5.3.4.1   AARP Post-Study Oral Assessment Intervention exercise (IE) results  212
    5.3.4.2   AARP Post-Study Oral Assignment assessment results  213
  5.3.5   AARP Post-Study Intervention Questionnaire results  217
  5.3.6   AARP Post-Study - Overview of outcome  219
5.4       Rating frequency ratios throughout the AARP, 2005-2010  222
  5.4.1   Rating frequency ratios for the AARP 1st Writing Assignment assessment, 2005-2010  223
  5.4.2   Rating frequency ratios for the AARP 2nd Writing Assignment assessment, 2005-2010  226
  5.4.3   Rating frequency ratios for AARP Oral Assignment assessment, 2005-2010  229
5.5       Concluding Remarks  230
CHAPTER 6 DISCUSSION OF AARP RESULTS

6.0 Introduction 232
6.1 Writing assessment in the AARP: Discussion and analysis 233
   6.1.1 Outcomes of Assessment in the LM 1st Home Writing Assignment 233
   6.1.2 Outcomes of Assessment in the 2nd Home Writing Assignment 236
   6.1.3 Special cases and phenomena in AARP writing assessment 239
       6.1.3.1 Self-assessment maturity 240
       6.1.3.2 The unique success of the Pre-Study 242
       6.1.3.3 Variables within large groups 244
       6.1.3.4 Rogue groups 247
       6.1.3.5 The consequences of writing assessment intervention in the Post-Study 250
6.2 Oral assessment in the AARP: Discussion and analysis 251
   6.2.1 Special cases and phenomena in AARP oral assessment 255
       6.2.1.1 The consequences of oral assessment intervention in the Post-Study 256
       6.2.1.2 The rogue group 257
       6.2.1.3 Justifying comments for grades awarded 260
6.3 Answers to the AARP research questions 262
6.4 Concluding remarks 267

CHAPTER 7 CONCLUSION

7.0 Introduction 268
7.1 The aims of the AARP and whether they were achieved 268
7.2 The success of the AARP 270
   7.2.1 Replication of research conditions 271
   7.2.2 The missing autonomy link 271
   7.2.3 Cross-sectional and longitudinal perspectives 272
7.3 Usefulness and transferability of the research 273
   7.3.1 The AARP model 273
   7.3.2 The AARP ‘glocalised’ definition of autonomy 274
   7.3.3 The AARP quantitative and qualitative data 274
   7.3.4 Two types of peer assessment 275
   7.3.5 Use of a single instructor-rater in the AARP 276
   7.3.6 The simplicity of the AARP 278
7.4 Limitations of the AARP 279
   7.4.1 Assessment feedback on the AARP 279
   7.4.2 Individual assessment behaviour 282
   7.4.3 An examination of outliers 283
   7.4.4 Lack of longterm development 284
7.5 Conclusions and recommendations 284
   7.5.1 Preparation of learners 285
   7.5.2 Preparation of teachers 286
   7.5.3 Use of technology 287
7.6 Final concluding remarks 288
Bibliography 290
<table>
<thead>
<tr>
<th>Figure No.</th>
<th>Description</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Nolen’s view of ‘intellectual’ autonomy</td>
<td>19</td>
</tr>
<tr>
<td>2.2</td>
<td>Combining Nolen and Kumaravadivelu’s views</td>
<td>21</td>
</tr>
<tr>
<td>2.3</td>
<td>The autonomy continuum</td>
<td>22</td>
</tr>
<tr>
<td>2.4</td>
<td>Mid-way on the heteronomy - autonomy continuum</td>
<td>23</td>
</tr>
<tr>
<td>2.5</td>
<td>Products vs. Producers of society</td>
<td>28</td>
</tr>
<tr>
<td>2.6</td>
<td>The heteronomy - autonomy continuum with Huttunen’s 3 levels of reflection</td>
<td>30</td>
</tr>
<tr>
<td>2.7</td>
<td>A learning continuum in which banking/ownership plays a key role</td>
<td>33</td>
</tr>
<tr>
<td>2.8</td>
<td>A continuum of autonomy with increasing self-determination</td>
<td>36</td>
</tr>
<tr>
<td>2.9</td>
<td>The expanded heteronomy - autonomy continuum</td>
<td>38</td>
</tr>
<tr>
<td>2.10</td>
<td>Three models for autonomous learning offered by van Lier (1990), Littlewood (1996; 1997) and Kohonen (1998)</td>
<td>57</td>
</tr>
<tr>
<td>2.11</td>
<td>Little’s 4 levels of necessary understanding for autonomy</td>
<td>59</td>
</tr>
<tr>
<td>2.12</td>
<td>A continuum of assessment offering degrees of autonomy</td>
<td>60</td>
</tr>
<tr>
<td>3.1</td>
<td>Reflection and assessment on the heteronomy-autonomy continuum</td>
<td>83</td>
</tr>
<tr>
<td>3.2</td>
<td>Stages of reflection and assessment on the heteronomy-autonomy continuum</td>
<td>96</td>
</tr>
<tr>
<td>3.3</td>
<td>Stages of apprenticeship on the heteronomy-autonomy continuum</td>
<td>100</td>
</tr>
<tr>
<td>4.1</td>
<td>Assessment Cycle for Writing Skills and Oral Skills</td>
<td>136</td>
</tr>
<tr>
<td>4.2</td>
<td>The LM I Writing Assessment Cycle, with feedback</td>
<td>146</td>
</tr>
<tr>
<td>5.1</td>
<td>y-graph of 1st Writing Assignment rating behaviour (2005-2010)</td>
<td>225</td>
</tr>
<tr>
<td>5.2</td>
<td>y-graph of 2nd Writing Assignment rating behaviour (2005-2010)</td>
<td>227</td>
</tr>
<tr>
<td>5.3</td>
<td>y-graph of Oral Assignment rating behaviour (2005-2010)</td>
<td>230</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>18</td>
</tr>
<tr>
<td>2.2</td>
<td>24</td>
</tr>
<tr>
<td>2.3</td>
<td>25</td>
</tr>
<tr>
<td>2.4</td>
<td>32</td>
</tr>
<tr>
<td>2.5</td>
<td>40</td>
</tr>
<tr>
<td>2.6</td>
<td>51</td>
</tr>
<tr>
<td>2.7</td>
<td>55</td>
</tr>
<tr>
<td>2.8</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>134</td>
</tr>
<tr>
<td>4.2</td>
<td>135</td>
</tr>
<tr>
<td>4.3</td>
<td>137</td>
</tr>
<tr>
<td>4.4</td>
<td>139</td>
</tr>
<tr>
<td>5.1</td>
<td>168</td>
</tr>
<tr>
<td>5.2</td>
<td>170</td>
</tr>
<tr>
<td>5.3</td>
<td>171</td>
</tr>
<tr>
<td>5.4</td>
<td>173</td>
</tr>
<tr>
<td>5.5</td>
<td>175</td>
</tr>
<tr>
<td>5.6</td>
<td>180</td>
</tr>
<tr>
<td>5.7</td>
<td>182</td>
</tr>
<tr>
<td>5.8</td>
<td>183</td>
</tr>
<tr>
<td>5.9</td>
<td>184</td>
</tr>
<tr>
<td>5.10</td>
<td>187</td>
</tr>
<tr>
<td>5.11</td>
<td>188</td>
</tr>
<tr>
<td>5.12</td>
<td>190</td>
</tr>
<tr>
<td>5.13</td>
<td>191</td>
</tr>
<tr>
<td>5.14</td>
<td>193</td>
</tr>
<tr>
<td>5.15</td>
<td>195</td>
</tr>
<tr>
<td>5.16</td>
<td>197</td>
</tr>
<tr>
<td>5.17</td>
<td>198</td>
</tr>
</tbody>
</table>
5.18 Mean values of Group G and H (2008-2009) questionnaire responses and independent t-test results

5.19 AARP Post-Study, 2009-2010 - Group Breakdown

5.20 AARP Post-Study Values for 1st Writing Assessment Intervention

5.21 AARP Post-Study 1st Writing Assignment Mean Scores and ANOVA results

5.22 AARP Post-Study 1st Writing Assignment assessment – Tukey-Kramer mean comparisons and Pearson correlation coefficients

5.23 AARP Post-Study 2nd Writing Assignment Mean Scores and ANOVA results

5.24 AARP Post-Study 2nd Writing Assignment – Tukey-Kramer mean comparisons and Pearson correlation coefficients – ‘pure’ and ‘tainted’ groups compared

5.25 Post-Study Oral Assessment Intervention results

5.26 AARP Post-Study Oral Assignment Mean Scores and ANOVA results with all participants

5.27 AARP Post-Study Oral Assignment Mean Scores and ANOVA results with Intervention participants only

5.28 AARP Post-Study Oral assignment assessment results – Tukey-Kramer mean comparisons and Pearson correlation coefficients - Mixed, and Intervention participants only

5.29 Mean values of Group I and J (2009-2010) questionnaire responses and independent t-test results

5.30 Rating frequency ratios for the 1st Writing Assignment

5.31 Rating frequency ratios for the 2nd Writing Assignment

5.32 Rating frequency ratios for Oral Assignment assessment
Chapter 1
Introduction to the thesis

1.0 Introduction

The title of this thesis, ‘Degrees of Autonomy in Foreign Language Learning’, indicates the pivotal point on which the investigations and work which form this thesis are based; that is to say, that there are degrees of autonomy\(^1\) which can be achieved in the language-learning environment (Benson, 2001; Broady & Kenning, 1996; Candy, 1991; Gibbs, 1979; Grenfell & Harris, 1999; Henner-Stanchina, 1976; Holec, 1981; Hurd, 1998a; 1998b; Huttunen, 1990; Karlsson, Kjisik & Nordlund, 1997; Kohonen, 1992a; Kumaravadivelu, 2003; Little, 1995a; McDevitt, 1997; Nunan, 1997; O’Leary, 2007a; Page, 1992; Paiva & Braga, 2008; Sinclair, 1997; 2000; Thanasopoulos, 2000), which in this instance is in a higher education (HE) setting.

The purpose of this chapter is first of all to provide some background to the Assessment for Autonomy Research Project (AARP), which was conducted in the context of the School of English (SOE), Aristotle University (AUTH) during the academic years 2005-2010. The significance of the study and the two preliminary research questions on which it was based will be presented, as will the main aims and scope of the study and the four main research questions which the study set out to investigate over a five-year period.

1.1 The backdrop to the AARP

The importance of autonomy in foreign language learning (FLL) has been highlighted by researchers for more than three decades, with all of them viewing autonomy not

---

\(^1\) For the purposes of this thesis, autonomy will be treated as synonymous with both independent learning and self-directed learning, and will be examined in relation to heteronomy. No definition is provided here, since one of the aims of Chapter 2 is to find a definition of autonomy which is appropriate for the context of the AARP.
only as being necessary in the context of language learning, but as being more widely applicable to using a language, to learning in general and to life itself. Little (1995b, p. 13) regards it as a “fundamental behavioural capacity”, while, similarly, Noels (2009, p. 302) sees it as a “human propensity”. Henner-Stanchina and Riley (1978, p. 75) view autonomy as “a philosophy of learning”, while Breen and Mann (1997, p. 134) posit that it is something much more fundamental in that it is “a way of being in the world” (authors’ emphasis), while, similarly, Brockett and Hiemstra (1991, p. 19) go so far as to say that it is “a way of life”.

These views are of particular relevance to this study, given that its participants are young adults in an HE setting. McNair (1997, cited in Hughes, 2003, p. 4) stresses that students arrive in HE with very “diverse” levels of autonomy and if they continue to learn in the same ways as they did previously, “there is a danger that individuals will become less, rather than more, autonomous”.

Barnett (2007, p. 29) believes that in addition to offering high levels of knowledge, understanding and skills, that it is the role of HE “to foster the development of human qualities and dispositions, of certain modes of being appropriate to the twenty-first century”. Although Barnett is more concerned with what he refers to as “authenticity” rather than autonomy, it is clear that the “dispositions” and the “modes of being” that he refers to equate with those of Benson (2009, p. 18), as well as Breen and Mann (1997, p. 134), who are concerned with autonomy, and Barnett also makes it clear that he believes that those dispositions and modes of being can be achieved through particular approaches to assessment. It is interesting to note that many researchers in the field of autonomy stress this essential link with assessment, with Dam (1995, p. 49) referring to it as the “pivot”, Harris (1997, p. 12), the “pillar”, and Little (2009a, p. 2), the “hinge” on which autonomy
depends. Hunt, Gow and Barnes (1989, p. 207) even go so far as to say that without self-assessment “there can be no autonomy”, a view which is shared by the present Instructor-Researcher (I-R) and which will be investigated in this thesis.

This thesis therefore investigates the key role which assessment has to play in the engendering of autonomy and how, as Harris & Bell (1990, p. 111) assert, the greater the degree of learner engagement with assessment, which in the case of the AARP is concerned with the assessment of language learning, the greater the degree of autonomy that can be exercised.

1.2 The significance of the AARP

Although English as a Foreign Language (EFL) theoreticians and experts in the field of autonomy in language learning (ALL), such as Dam (1995), Dam and Little (1998), Dickinson (1987; 1992), Little (2007a; 2007b), Nunan (1988), Oscarson (1989; 1997) and Oskarsson (1978; 1984), have delivered a very strong and clear message concerning the important role that assessment has to play in the fostering of autonomy, nevertheless, in recent EFL studies which have emerged, which were concerned with the exploitation of peer- and self-assessment of speaking or writing skills, the possibility of a link between autonomy and assessment has, for the most part, either been mentioned in passing or has been completely ignored. It was, therefore, out of a desire to further explore the ambivalent relationship between assessment and autonomy in FLL that the AARP was developed.

The AARP was a five-year study undertaken in relation to autonomy in FLL, which, in this instance, was EFL. It was implemented in the SOE of AUTh and involved the investigation of peer-assessment and self-assessment of writing and

---

2 The words ‘assessment’ and ‘evaluation’ can be used interchangeably, but the former is generally preferred in the context of ‘self’ and ‘peer’ involvement.
speaking skills as a means to promote greater autonomy in language learning (ALL). The research study consisted of three phases extending over the five academic years 2005–2010: a Preliminary Study\(^3\) (2005-2006), the Main Study (2006-2009) and a Post-Study (2009-2010). The three study phases were conducted with students on the SOE 1\(^{st}\) Year course, Language Mastery I (LM I). Subjects used in the AARP study were members of the I-R’s own LM I course groups, which means that the cohorts used for research purposes were convenience samples.

Quantitative data concerning peer-, self- and instructor assessment of both writing skills and speaking skills was collected through criterial assessment checklists for two home writing assignments and one in-class oral assignment, in the form of a presentation, which were prepared at home. All the participants in the processes of assessment used the same criteria, facilitating triangulation of results and their analysis, while qualitative data was collected by means of assessment questionnaires, with responses given on a Likert scale, which could be analysed statistically, with some space given also to written comments, which could be analysed cognitively.

The AARP differs from all previous EFL assessment studies in four significant ways. Firstly, it was a replicative study, in that the same conditions were created, as far as possible, with two groups of students on the same course, in five consecutive years, replication being something which researchers such as Boud and Falchikov (1989) complain is missing from the assessment literature generally. Unlike the AARP, all research studies previously conducted in an FL or an EFL HE context have had a duration of one semester only.

Secondly, in the AARP, peer- and self-assessment processes were applied to the two skills of speaking and writing. With the exception of Cheng and Warren (1999;

\(^3\) The AARP Preliminary Study will henceforth be referred to as the Pre-Study.
2005), peer-assessment and self-assessment have been conducted either in relation to speaking skills or to writing skills, but not in relation to both skills with the same group(s) of students. Having records of student assessment behaviour regarding the two productive skills, as in the case of the AARP, offers very interesting results for comparison, and, in the case of some cohorts, confirmation of assessment abilities.

Thirdly, the type of discourse or genre with which the LM I course and the AARP were concerned was that of description and narrative. While the latter is not unusual in the case of studies concerned with the assessment of speaking skills, it is the exception rather than the rule in the case of studies concerned with the assessment of writing skills, which, as a whole, deal with discursive or persuasive essay writing.

Fourthly, although most studies provide for some kind of qualitative feedback from students concerning assessment processes, because the studies were of short duration and in most cases were conducted with small samples, there is nothing on a similar scale to the feedback data which was gathered during the five-year period when the AARP was conducted. The data which has been gathered from participants in the AARP is a truly rich resource, offering a window on student insights and perceptions. Some examples of these perceptions, gleaned from replies to questionnaires, are offered in Appendices 26, 27 & 28.

1.3 Preliminary research questions
Before embarking on the present research study, it was necessary to look at existing models in the autonomy literature to see if any would be suitable for adoption for the AARP; however, in order to choose a suitable model, it was also essential to try and establish exactly which conceptualisation of autonomy the model would be based upon. This required that a definition of autonomy, on which the whole thesis would be
based, be outlined. The two preliminary research questions which therefore needed to be answered were:

(i) Can a definition of autonomy be found, on which the AARP model and the practices within could be based, which would be suitable for the context of the SOE, AUTh?

(ii) Can a suitable model for autonomy in foreign language learning (FLL) be found which would allow for varying and fluctuating degrees of autonomy?

The research as a whole is based on the working hypothesis that by adopting a model for language learning which allows for variation and fluctuation in language learner behaviour, on a spectrum or continuum between heteronomy and autonomy, and by accepting a definition of autonomy which permits of variation in disposition and uptake, it is possible to allow for learner differences and view engagement with autonomy in language learning as varying in degrees. Such a model would have to take into account the four questions which Little (1996a, p. 3; 1999a, p. 28) views as being essential to autonomy in language learning:

(i) What are we learning?
(ii) Why are we learning?
(iii) How are we learning?
(iv) With what success?

It is clear from the last question, in particular, the important role that assessment has to play in helping learners move from a state of heteronomy, where they are dependent on others within the educational process for decision-making and direction, to one of autonomy, where more of the decisions, directions and initiatives are taken by the learners themselves.
1.4 The aims and scope of the research

The research conducted in the AARP is based on several premises which have been posited by experts in the field. Holec (1985a) asserts that autonomy is a conceptual tool and must be used as such. This is important in the present context as one of the aims of the AARP was to develop critical thinking skills through involvement in peer- and self-assessment processes. Grenfell and Harris (1999) believe that good language learning is concerned with increasing autonomy and they, together with Nunan (1997) and Little (1995c; 2007b) insist that autonomy is also a prerequisite in effective language use, creating, in turn, competences which are conducive to lifelong learning (Little, 1995b; Macaro, 1997; Stefani, 1998). Likewise, Hurd (1998a, p. 220; 1999, p. 7) believes that in order for there to be any hope of achieving autonomy, that the onus lies equally well with the teacher and the learner to actively seek out what amount to “good learning opportunities”. It was hoped, in the AARP, that through involving learners in the assessment process that such “opportunities” for learning, for exercising autonomy and for practising lifelong learning skills would be created.

Having sought satisfactory solutions to the two preliminary research questions of finding a suitable definition and model for the AARP, the next course of action was to investigate the four main research questions in the AARP, which aimed to establish the following, in an EFL higher education context:

(1) Can learners assess the oral and writing skills of their peers (peer-assessment), according to predetermined criteria with objectivity and reliability?

(2) Using peer-assessment as a stepping-stone, can learners then assess their own oral and writing skills performance (self-assessment) with objectivity and reliability, using the same predetermined criteria?
(3) Is there evidence that learners can assume ownership of the assessment criteria checklists and exercise judgement in an atmosphere of cooperation and trust?

(4) Is there evidence that the cyclical process of peer- (P-A), self- (S-A) and instructor-assessment (I-A) can combat pre-conceptions about assessment, turning it from a possible constraint to learning into an affordance and thus enable learners to use it as a learning opportunity which might lead to greater self-direction and autonomy?

Macaro (1997, p. 168) suggests, as do Holec (1981; 1985a; 1985b) and Hunt, Gow and Barnes (1989), that inherent in the concept of learner autonomy is the learners’ “knowing how to make decisions” in addition to “being allowed to make those decisions” (author’s emphasis). It was hoped that the procedures and practices conducted in the AARP would help shape learner thinking and behaviour, through reflection and decision-making, and in this way lead them to greater autonomy. Indeed, the fact that greater autonomy may be offered to learners, is no guarantee of its uptake. Like the previously-mentioned McNair (1997, cited in Hughes, 2003), Ridley (1997, p. 18) reminds us that “learners of all ages and in all learning situations are capable of not acquiring the various types of metacognitive/self-regulating skills which underpin autonomous learning” (author’s emphasis). Clearly, then, the promotion of autonomy in the language classroom, even at the level of tertiary education, presents a challenge both to the learners and their instructors. The AARP was an attempt to meet that challenge. In Chapter 2, Section 2.3, some of the elements which appear to “underpin” autonomous learning and to be essential to its development, are examined.
1.5 Organisation of the thesis

The thesis consists of two volumes. In the first volume, there are a total of seven chapters, while the second volume, which will be referred to as Volume 2 (Vol. 2), contains all the appendices, grouped thematically into seven sections.

In the first volume, Chapter 2 examines the literature on autonomy which spans over the past three decades and through the exploration of areas of consensus regarding autonomy amongst theoreticians hopes to answer the first research pre-question and find a definition which suits the context of the AARP. Various, seemingly disparate conceptualisations and models which have been proposed for autonomy by a number of experts from the field of applied linguistics and from other disciplines will also be scrutinised and compared in an attempt to find common elements which will lead to a unified model which seems appropriate for the AARP. In this way, it is hoped to answer the second research pre-question.

Chapter 3 offers an overview of the assessment literature which relates to autonomy and provides some background to the present study, with particular emphasis being placed on relevant work within the Council of Europe. It will also briefly focus on the work of two world-leading proponents of self-assessment and peer-assessment: Boud and Falchikov. This chapter will conclude with an overview of the studies in the EFL literature most relevant to the AARP.

Chapter 4 outlines the pedagogical procedures and research methodology, which include the teaching approach, the teaching materials used, the instruments used and the way in which assessment was conducted and how the results of assessment and the opinion of AARP participants were recorded and later analysed. Any variations in procedures will be highlighted and explained.
Chapter 5 presents the results from the three stages of the AARP, namely the Pre-Study, the Main Study and the Post-Study and will conclude with a brief synoptic overview of assessment rating behaviour for the five-year duration of the study.

A discussion of the results of the AARP follows in Chapter 6, in which answers to the four main research questions are sought and interesting cases which emerge from the study will be examined in more detail, with reference to other appropriate studies as well as quantitative and qualitative data from the AARP.

The final chapter, Chapter 7, distilling the information from the previous chapters will examine the overall success of the AARP and what it has contributed to knowledge in the area. Possible limitations and weak points will be pinpointed and suggestions will be made concerning possible areas worthy of future exploration.

This draws the thesis to its conclusion.

1.6 Concluding remarks

In this introductory chapter the importance of autonomy in language learning and in life and the significance of the AARP were underlined. The two preliminary research questions concerning the finding of a suitable definition and the finding of a model appropriate to the context of the AARP were posited. The aims and the scope of the research and the four research questions to be examined were outlined. Finally, the way in which the remaining six chapters in the thesis are organised was elaborated upon.
Chapter 2
Autonomy in language learning: an overview

2.0 Introduction
This chapter will provide an overview of expert opinion, in the ALL literature, of relevance to the AARP. Firstly, the chapter will examine why autonomy is of interest to us in language teaching and the implications of promoting autonomy. Through analysis, an attempt will be made to synthesise the various theories suggested for ALL, within a proposed continuum of autonomy vs. heteronomy. Some key, but somewhat neglected elements in the promotion of autonomy will be discussed briefly, making more comprehensible the role played by those same elements in a model for autonomy. Finally, an attempt will be made to draw together, with supporting documentation, what seem to be 30 areas of consensus concerning what autonomy, in an FLL context, is. This will greatly facilitate the formulation of a definition of autonomy appropriate to the AARP and will likewise help lay the foundations for a suitable AARP model, with which the chapter ends.


It is because of the very perplexity surrounding autonomy that the aim in this chapter will be to analyse and synthesise, to identify areas of commonality amongst what can appear to be very diverse views and by drawing these various threads of
commonality together, show the very fabric from which an abstract concept such as autonomy is made.

For this reason, one of the aims in this chapter will be to look at existing definitions of autonomy in an EFL context and decide on a definition which would be suitable for the context of the AARP. This will therefore be an attempt to answer the first pre-research question, already given in Chapter 1, Section 1.3, but repeated here for convenience:

(1) Can a definition of autonomy be found, on which the AARP model and the practices within could be based, which would be suitable for the context of the SOE, AUTh?

Having attempted to find a definition, the second aim will be to examine existing models of autonomy and self-direction which have been presented by experts and through this examination to find our own model for autonomy which encapsulates both the theory and the practice behind the AARP. There will be an attempt to answer the second research pre-question, presented in Chapter 1, Section 1.3, and repeated here, which is:

(2) Can a suitable model for autonomy in foreign language learning (FLL) be found which would allow for varying and fluctuating degrees of autonomy?

Before searching for a suitable definition and model for the AARP, it will be necessary to examine the concept of autonomy in more detail.

2.1 Why autonomy?

Given the complexity of autonomy as a concept (Paiva & Braga, 2008), and given the even greater complexity of trying to put autonomy into practice in the language
classroom, and taking into account the possibility that learners may not really wish to have autonomy conferred on them at all (Little, 1990), it is perhaps important to consider why, as language teachers, we should be concerned with autonomy and its promotion in the first place.

Autonomy, in the sphere of education, is a concept which has attracted great interest and has also created controversy, but within the sphere of language education, autonomy holds still greater appeal, because besides the empowerment of language learners which autonomy seems to offer, both in learning and using the language (Little, 1995c; 1997; 2010; Grenfell & Harris, 1999), the whole process of becoming autonomous as a language learner would seem to involve the acquiring of dispositions and behaviours4 which can be applied not only in the mastery of other languages, but can be applied to learning in other domains (Little, 1995b) and not just through primary, secondary and tertiary education, but throughout a person’s life (Third Nordic Workshop, 1990; Little, 1995b; Hughes, 2003). In this way, individuals will not only have become autonomous learners and users of a second, and possibly a third or more languages, but they will have acquired skills enabling them “to take some control over their lives” (Clifford, 1999, p. 115) and will also be on their way to becoming lifelong learners (Clifford, 1999; Falchikov, 2007), something which is deemed a prerequisite for survival in the modern age (Longworth & Davies, 1996), with its “supercomplexity” (Barnett, 2000, cited in Barnett, 2007, p. 29). Pekkanli Egel (2009, p. 2023) sees it as the duty of language teachers to become aware of the factors conducive to learner autonomy since it is the “key” to lifelong learning.

4 Benson (2009, p. 18) states that if it is possible to define autonomy, “it must be defined as a composite of abilities, attitudes or dispositions” (his emphasis). Parrott (1993, p.1, cited in Allwright & Hanks, 2009, pp. 69-70), very importantly, also reminds us of the range of predispositions to learning within any one group of learners. Kember & Gow (1989) and Fazey & Fazey (2001) offer detailed discussion about learner predispositions to learning and autonomy.
Longworth & Davies (1996, p. 39) point out that the complexities of modern-day living demand that in order to survive, everyone must be capable of handling information “quickly and adaptably”, something that the skills, acquired in school and continued at university, with learning at a surface level of simply memorizing data and rewriting them in exam conditions, has not enabled most people to do.

Marshall (1996, p. 40), like Stefani (1998, p. 349), believes that the job of higher education should be to ensure that undergraduates achieve their full and true potential, that the aim should be to achieve “intellectual growth” so that they are able to cope effectively with the “world of work”, are also able to “contribute to society” and reach the maximum degree of “personal fulfillment”. They also take the view that in order for young people in higher education to be prepared for survival in the global employment market that they need to be prepared by their instructors to take a much more pro-active stance to learning and to be able to set their own goals and targets. The qualities that Stefani sees learners in higher education as needing to acquire are those of flexibility, adaptability and the ability to take responsibility for their own continuing “personal and professional development” and she sees this as being achievable through the forming of partnerships between learners and their instructors. Stefani (1998, pp. 339-340) sees assessment as playing a key role in the achievement of autonomy and in preparation for lifelong learning.

Likewise, Little (1997, p. 35) views autonomy as the concept which above all others defines “the desired relation between formal learning and living” and helps to remove the barriers which can exist between the two (Little, 1999c, p. 11). However, Little (1995b, p. 13) feels that too much emphasis has been placed on what autonomy offers in terms of independence and freedom from constraint, while more attention,
he thinks, should be paid to autonomy as a “fundamental behavioural capacity” (Little, 1995b, p. 13), previously mentioned in Chapter 1, Section 1.1. In more recent years, he and other experts in the field have been giving emphasis to the social, cooperative and collaborative aspects of autonomy, which involves working together with others as members of a learning community in a state of “relatedness” (Littlewood, 1999; Hughes, 2003; Little, 2007a; Carson, 2007; Smith & Ushioda, 2009). Hence, one of the keys to our understanding of learner autonomy is the ability to comprehend that autonomy is “the product of interdependence rather than independence” (Allwright, 1990; Little, 1994; Benson, 2001).

2.1.1 The implications of promoting autonomy
According to Coyle (2000, p. 24-25), such an intention to empower individuals “to take control of the language they are learning” makes tremendous demands on teachers, whom he believes have inherited “a fossilised orthodoxy of practice” left behind by Communicative Language Teaching (CLT), and also makes great demands on learners, whom he believes, like Legutke and Thomas (1991) take a “reactive” rather than an “interactive” role in the language-learning process and suffer from what Legutke and Thomas (1991, p. 7) refer to as the “dead bodies, talking heads” syndrome, which, according to Coyle (ibid) is characterized by their being “disaffected, demotivated and subsequently linguistically incompetent adolescents”.

Allwright and Hanks (2009, p. 48) are also highly critical of CLT because it has done “little”, in their opinion, to promote the view of the learner as a “developing practitioner”, which they believe is so essential in the fostering of autonomy. While these may seem like pessimistic views of how things stand with regard to language learning and teaching and do not seem to bode well for the outcome of the present
project, it is true that in our consideration of learners and in our desire as teachers to promote autonomy in our learners, account has to be taken of the constrictions imposed by previous experiences of learning and of assessment (Ecclestone & Pryor, 2003; Ecclestone, 2002), which may have been reinforced over a number of years.

There are also the effects of parental, peer and societal pressure, as well as the pervading attitudes to learning and education that exist within the departmental student population in the SOE, in addition to other limitations, restrictions and constraints, and sometimes, antithetically, the lack of them, concerning time, curriculum, assessment procedures and course attendance imposed by university regulations and education laws.

With regard to language instruction, and language instructors in particular, Little (1994, p. 441) makes the point that “learner autonomy imposes new and heavy demands on teachers”. Edge (1997, p. 27) also believes that such a move towards autonomy “makes enormous demands on teachers”. Strevens (1980, p. 17) reminds us that “it takes better teachers to focus on the learners”, unfortunately without really elucidating on what qualities might be required of the teacher to make them “better” and facilitate that focusing; however, some idea of the kind of qualities necessary and of the demands that might be placed on teachers in a more learner-centred classroom are outlined for us by Breen & Mann (1997, p. 145), who list the three “attributes” necessary in teachers who wish to promote autonomy, as follows:

(1) Self-awareness, in that they understand the situation in which learners find themselves and are aware that the teachers’ decision-making and behaviour, and the behaviour they demand of their students, will ‘influence that autonomy’.
(2) There must be trust and teachers must believe that learners are capable of taking autonomy on board.

(3) There must be a true desire on the part of teachers to promote autonomy and a willingness to accept whatever results from that.

In addition to this, Breen & Mann (1997, pp. 146-148) also visualize six ways in which teachers can facilitate and “may create space for learners to exercise their autonomy” and these are:

(1) Acting as a resource
(2) Permitting learners to help with decision-making
(3) Promoting collaborative assessment
(4) Coping with risk-taking
(5) Exploiting opportunities as they present themselves
(6) Accepting support.

In relation to the training of teachers who wish to pursue a more learner-centred approach to teaching, where they are “no longer expected to tightly orchestrate and dominate all work in the classroom”, Murdoch (1990, p.15) reminds us that training or retraining teachers, particularly experienced ones, for this kind of classroom set-up can pose “intractable problems”.

Despite the heavy demands which promoting autonomy make on the teacher, nevertheless, many practitioners and researchers take the view that the empowerment of language-learners (Coyle, 2000, p. 36) is an ideal worth pursuing and is an endeavour, no matter how difficult and demanding, that should be undertaken and encouraged. Little (1996a, p. 2) suggests that autonomy is not something which will occur accidentally, but “must necessarily be a matter of conscious intention”.
2.2 Autonomy vs. heteronomy

While the derivation and etymology of the word ‘autonomy’ is frequently referred to in the autonomy literature, the fact that this word has an opposite, both in Greek and in English (see Table 2.1 below), is, for the most part, ignored. This appears to be something of an oversight, as it seems clear that any move towards the promotion of autonomy in language learning is, by its nature, a move away from the promulgation of heteronomy.

Table 2.1 Derivation and meaning of autonomy and heteronomy

<table>
<thead>
<tr>
<th>WORD</th>
<th>GREEK DERIVATION</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>autonomy</td>
<td>αυτός = self</td>
<td>government, rule, regulation, direction of self</td>
</tr>
<tr>
<td></td>
<td>νόμος = law</td>
<td></td>
</tr>
<tr>
<td>heteronomy</td>
<td>έτερος = other</td>
<td>government, rule, regulation, direction by others</td>
</tr>
<tr>
<td></td>
<td>νόμος = law</td>
<td></td>
</tr>
</tbody>
</table>

Surprisingly enough, while educational psychologists such as Nolen (1995) and Waite-Stupiansky (1997) make clear the autonomy/heteronomy distinction, the word heteronomy, with few exceptions (Namenwirth, 1996; Rujiketgumjorn, 2000; Schmenk, 2006; 2008), rarely appears in the language education literature. Even though this distinction and awareness of the bipolarity between autonomy and heteronomy is not usually stated, it could be that an unconscious awareness of the alternatives to autonomy leads researchers such as Wright (2005, p. 134) to regard autonomy as an “entitlement”, while others such as Hustler and Hodkinson (1996, p. 111) and Rujiketgumjorn (2000, p. 2) suggest it is a “right”.

Indeed, perhaps it is imperative that we should consider autonomy as a right, particularly if we consider the alternatives to autonomy. According to the educational
psychologist, Nolen (1995, p. 202) (see Figure 2.1 below), if what she terms “intellectual autonomy” is not attained during our years of schooling and membership of the academy, she fears that we are in danger of remaining in a state of “intellectual heteronomy” where we are forever dependent on more knowledgeable “others” to inform us of what is true or untrue and whether and how much we have been successful in our understanding and learning. Boud (1981, p. 13) also shares Nolen’s fears that without the ability to judge and assess our learning achievements, we remain dependent on others to be the judges of our “competence”. Waite-Stupiansky (1997, p. 23), indeed, claims that schools have traditionally tended to promote heteronomy more than autonomy. Grow (1991, p. 129) points out that such dependency is not a “defect”, but can be a “serious limitation”, while Schmenk (2006, p. 81) stresses the importance of understanding the “dialectics” between autonomy and heteronomy and not regarding them as two extremes of “good” and “bad”. Little (2000c, p. 9) regards the two as having a “symbiotic relation”, with the need and opportunities for support, from the teacher and peers, varying according to the task in hand and the stage of learning.

![Figure 2.1 Nolen’s view of ‘intellectual’ autonomy](image)

Barnes (1976, as cited in Dam & Little, 1998) takes a similar view to Nolen’s, but with a slightly different slant. He makes the useful distinction between what he terms “school knowledge”, which is presented by “someone else” and “action knowledge”, on which learners’ actions and choices are based, as a result of
“reflection”. Bruner (1986, as cited in Dam & Little, 1998), sees knowledge of the type described by Barnes as “school knowledge” as having control over the learner, “from the outside in”, while learners who have become capable of what he calls “reflective intervention”, are able to function “from the inside out” and “will control and select knowledge as needed”. Schelfout, Dochy and Janssens (2004, p. 179), in talking about the learning environment in the higher education sector, seem to replace Barnes’s term “school knowledge” with that of “inert knowledge”, which they believe results from “traditional lecture-based instruction”.

Little (1991, p. 13) thinks that still in the popular imagination, learning is viewed as the assimilation of a body of facts, which is “static” and regards the educational psychologist and humanist Rogers (1967, cited in Little, 1991) as being one of the first to highlight the need for individuals to be able to cope with “rapid changes in knowledge”. Little (1990, p. 13) makes it clear to us that learning is not just a simple matter of learners adding new knowledge to the knowledge which they have already acquired, but rather, for autonomous learners, it is something more fundamental, so that “they can integrate what they learn with the rest of what they are”.

In turn, the views of Barnes (ibid), Bruner (ibid) and Little (1990) seem to be closely connected to those of deCharms’s (1968, cited in Stefanou, Perencevich, DiCintio & Turner, 2004), of the learner either being an “origin” or a “pawn”. According to Stefanou et al. (2004, p. 98), while “origins” are able to set themselves targets and determine how to achieve them, “pawns”, on the other hand, feel themselves to be subject to external control and therefore “lack volitional strategies and behaviors”.
At the same time, it is interesting to take note of Kumaravadivelu’s (2003, p. 141) view on autonomy, where he makes a distinction between what he refers to as “academic autonomy” and “liberatory autonomy”. Kumaravadivelu sees “academic” autonomy as enabling learners to be “strategic practitioners” who are capable of realising their “potential”, while he sees “liberatory” autonomy as being more empowering so that learners become “critical thinkers” who are capable of realising their “human potential”. Thus empowered, learners are capable of recognising “impediments” to their learning and finding ways of circumventing them.

Thus, “liberatory autonomy” as outlined by Kumaravadivelu (ibid) seems to offer the same emancipatory qualities as the “intellectual autonomy” described by Nolen (1995), while the “academic autonomy” Kumaravadivelu speaks of would seem to belong somewhere along the spectrum or continuum between “intellectual heteronomy” and “intellectual autonomy”, so that a chart combining the two researchers’ views might look something like Figure 2.2 below:

![Figure 2.2 Combining Nolen and Kumaravadivelu’s views](image)

Despite the fact that heteronomy, as applied to language education, has been largely ignored, it is clear that many researchers such as Holec (1981), Huttunen (1986; 1990), Kohonen (1992a), Tait & Knight (1996), Rujiketgumjorn (2000) and Sinclair (2000) still think of autonomy as being on some kind of continuum. Kohonen (1992a, p. 23), talks about the two extremes of the continuum as being “other-directed” and
“self-directed”, while Sinclair (2000, p. 8) refers to the “vegetative” state of “complete lack of autonomy” as compared with the “idealistic” extreme of “complete autonomy”, whereas Huttunen (ibid) distinguishes between “teacher-centred” and “learner-centred” language education. All of these views would seem to differ only in the terminology they use and can be combined, as in Figure 2.3 below:

<table>
<thead>
<tr>
<th>HETERONOMY</th>
<th>AUTONOMY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nolen, 1995; Namenwirth, 1996; Waite-Stupiansky, 1997; Rujiketgumjorn, 2000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OTHER-DIRECTED</th>
<th>SELF-DIRECTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holec, 1981; 1985a Kohonen, 1992a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TEACHER-DIRECTED</th>
<th>STUDENT-DIRECTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huttunen, 1986; 1990</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COMPLETE LACK OF AUTONOMY (VEGETATIVE)</th>
<th>COMPLETE AUTONOMY (IDEALISTIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinclair, 2000</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.3 The autonomy continuum**

Other researchers simply speak of “degrees of autonomy” without any reference to a continuum at all (Henner-Stanchina, 1976; Gibbs, 1979; Holec, 1981; Candy, 1991; Karlsson *et al.*, 1997; McDevitt, 1997; Hurd, 1998b; Grenfell & Harris, 1999; Benson, 2001; Kumaravadivelu, 2003). Holec (1985a, p. 2) does not speak of a continuum, but of a “progression” from being “other-directed” to “complete self-direction” (see Figure 2.3, above).

It would be an omission to ignore the other researchers from the field of autonomy in language learning who have offered different conceptualisations of the way that autonomy can vary in degree. These include Littlewood (1999, pp. 75-76), who talks of autonomy ranging from “reactive” to “proactive”, and Smith (2002, p.18), who views autonomy as ranging between what he refers to as a “weak” version and a “strong” version. It seems clear that in the case of Littlewood and Smith that
their concepts of “reactive” and “weak” autonomy, like Kumaravadivelu’s “academic” autonomy, lie somewhere between the two extremes of heteronomy and autonomy. Thus, if we add their conceptualisations of autonomy to the continuum, it will look something like Figure 2.4 below:

![Figure 2.4](image)

In the section which follows, we will turn our attention to some of the concepts and capacities which seem to be of significance in the promotion of autonomy, but which have not always been highlighted in the literature; nevertheless, they need to be taken into consideration.

### 2.3 Some key elements in the promotion of autonomy

Before moving on to the areas of consensus about autonomy and seeking a suitable definition, it is important that we have an understanding of some of the elements which underlie autonomy and which some experts consider require our attention if there is to be any hope of promoting autonomy. It seems clear that a progression from heteronomy towards autonomy, first of all, implies a moving away from traditional, transmissional and behaviourist models of education towards more transformative, Kolbian-experiential or Vygotskian, neo-constructivist models.
Something which will be useful in helping us to understand the implications, in all aspects of the pedagogical setting, of such a progression from behaviourist, transmissonal modes of learning to experiential/constructivist, transformative modes of learning is shown in Table 2.2 below, which is reproduced from Kohonen (1992a, p. 31):

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Traditional model: behaviorism</th>
<th>Experiential model: constructivism</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. View of learning</td>
<td>Transmission of knowledge</td>
<td>Transformation of knowledge</td>
</tr>
<tr>
<td>2. Power relation</td>
<td>Emphasis on teacher’s authority</td>
<td>Teacher as ‘a learner among learners’</td>
</tr>
<tr>
<td>3. Teacher’s role</td>
<td>Providing mainly frontal instruction; professionalism as individual autonomy</td>
<td>Facilitating learning (largely in small groups); collaborative professionalism</td>
</tr>
<tr>
<td>4. Learner’s role</td>
<td>Relatively passive recipient of information; mainly individual work</td>
<td>Active participation, largely in cooperative small groups</td>
</tr>
<tr>
<td>5. View of knowledge</td>
<td>Presented as ‘certain’; application, problem-solving</td>
<td>Construction of personal knowledge; identification of problems</td>
</tr>
<tr>
<td>6. View of curriculum</td>
<td>Static; hierarchical grading of subject-matter, predefined contents</td>
<td>Dynamic; looser organisation of subject-matter, including open parts and integration</td>
</tr>
<tr>
<td>7. Learning experiences</td>
<td>Knowledge of facts, concepts and skills; focus on content and product</td>
<td>Emphasis on process: learning skills, self-inquiry, social and communication skills</td>
</tr>
<tr>
<td>8. Control of process</td>
<td>Mainly teacher-structured learning</td>
<td>Emphasis on learner: self-directed learning</td>
</tr>
<tr>
<td>9. Motivation</td>
<td>Mainly extrinsic</td>
<td>Mainly intrinsic</td>
</tr>
</tbody>
</table>

From Kohonen (1992a)

Thus, a move from heteronomy towards autonomy in language learning implies a move from the model on the left-hand side of Table 2.2 to the model on the right-
hand side, with more control and responsibility for learning being offered to the learners. Another model, provided by Berry & Sahlberg (1996, p. 26), which is based on Brody (1991, cited in Berry & Sahlberg, 1996), provides us with a mid-way stage between heteronomy and autonomy, and thus a mid-way stage between Transmission and Transformation modes of learning (see Table 2.2 above), in that of Transaction (see Table 2.3 below):

**Table 2.3  Brody’s three typological stages of epistemological orientation**

<table>
<thead>
<tr>
<th>Epistemological orientation</th>
<th>TRANSMISSION</th>
<th>TRANSACTION</th>
<th>TRANSFORMATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The nature of knowledge and knowing</strong></td>
<td>Knowledge is transmitted from the teacher or text to the student. Knowledge is static and objective and knowing is seen as closed, linear paradigms. Quantity and breadth are emphasized.</td>
<td>Knowledge is gained through the interaction between the learner and his/her environment. Knowledge is dynamic and alive. Knowing is based on learning strategies. Quality of learning is emphasized instead of quantity.</td>
<td>Knowledge is dynamic, changing and it is constructed by the learner. Knowing is contextual, formal and informal discourse are essential when building the community of learners.</td>
</tr>
<tr>
<td><strong>Sense of authority</strong></td>
<td>Teacher-centred. Teaching is emphasized and the learner dependent on the teacher. Teacher is also responsible for the learning outcomes and the design of the learning environment.</td>
<td>Student-centred. Learners are responsible for their own learning with the teacher. Teacher has the control of the situation but he is not authoritarian. Strong intrinsic motivation is empowered.</td>
<td>Community of learners. Visible authority does not exist. Teacher uses the power of the environment and the community when new knowledge is created/transformed. Complexity, openness and creativity are emphasized.</td>
</tr>
<tr>
<td><strong>Conceptions of learning</strong></td>
<td>Learning is transferring knowledge and skills from teacher to learner. Effectiveness of learning is tested in achievement tests and mastery of the content is emphasized. Learning is understood as a linear and simple action.</td>
<td>Learning is empowered through cooperative activities, problem-solving and higher order thinking. Productive talk and positive interdependence among the students are essential characteristics of the learning process.</td>
<td>Learning is a change in learner’s experiences and values and the constructive, self-regulative and cooperative processes are emphasized. Learning is seen as a construction of the community of learners.</td>
</tr>
</tbody>
</table>

*Adapted from Berry & Sahlberg (1996)*

It becomes clear that Brody’s and Berry & Sahlberg’s Transmission and Transformation orientations correspond with Kohonen’s Traditional/Behaviourist and Experiential/Constructivist models (see Table 2.2), while the Transaction orientation
given in Table 2.3 above, lies somewhere between Kohonen’s Traditional-Experiential models and therefore lies somewhere mid-way on our heteronomy-autonomy continuum (see Figure 2.4 above), thus equating with Kumaravadivelu’s “academic” autonomy, Littlewood’s “reactive” autonomy and Smith’s “weak” autonomy.

In order to be able to comprehend how such a transition might be attempted or achieved, with a moving away from transmisssional/behaviourist/heteronomous orientations in learning, through transaction stages, towards more transformational/experiential/autonomous orientations, it is necessary to take a number of concepts into account.

The concepts of identity, reflection, ownership and self-determination are those which seem to lie at the heart of autonomy; however, they have tended to be scrutinised in isolation, or investigated by researchers not primarily concerned with autonomy in language learning. Due to their importance and their bearing, both on the AARP and on our search for a suitable definition and model for the present study, they will be examined briefly here.

2.3.1 Identity
Identity appears to be inherent in the idea of autonomy, since Breen & Mann (1997, p. 134) claim that autonomy implies a “robust sense of self”. Little (1999a, p. 31) reminds us that language learning “has profound implications for our behavioural potential” and that there is necessarily a strong link with “our sense of self, our sense of identity”. Doughty & Thornton (1973, p. 41) emphasise how closely and intimately our sense of being and identity are linked to our ability to use language, positing that “our view of the world is inseparable from the way we use language to shape it”.

26
While our language may shape our world, Imhoof (1991, p. 40) seems to take this idea still further since he believes that we reshape ourselves when learning a new language, and that:

(i)n order to become a language student, we must suspend our other selves temporarily and become primarily language learners. More than most subjects, language study creates a break in our normal existence. To put it more positively, it is a step towards reshaping our lives. We are creating a new person with a new language. …language learning requires the student to take on a new identity.

Indeed, Little & Dam (1998, p. 3) believe that in the language-classroom setting that learning is “anchored” in the identity which learners have “achieved” and through the collaborative and interactive processes occurring when learners “construct their shared learning space”. Like Imhoof, Brown (1973, p. 233) is also of the opinion that someone is “forced” to form a new identity in a second language, if they wish to become “competent” and feels that success in language learning is linked to our “ego” and to our “self-knowledge”, “self-esteem” and our “self-confidence”.

Clearly, we can see the importance of self-assessment here and understand that skilled practice in self-assessment can help in the development of a healthy “ego”. Kohonen (2001b, p. 32) also seems to concur with Brown (ibid) in that he believes that in order for autonomy to be developed, that the three qualities of self-awareness, self-understanding and personal effectiveness must be developed. At the same time, Cohen & Norst (1989) and Williams & Burden (1997) emphasise the fact that our “language” is very closely linked to our “identity” and that language is used “to convey this identity to other people” (Williams & Burden, 1997, p. 115), while, more seriously, “an attack on one is an attack on the other” (Cohen & Norst, 1989, p. 61).
Likewise, Brown (1973, pp. 233-234) reminds us that “communication implies a process of revealing oneself to another”, which is not easy.

The idea of ‘agency’ is very closely linked with identity and Benson (2006b, p. 30) sees the two as being “the key linking concepts between sociocultural theory and the theory of autonomy”. Little (1990), citing Ignatieff (1990), views autonomy as being an empowering force which enables learners to become the “artists of their own lives”, while Pennycook (1997) and Macaro (1997), likewise, see learners as becoming the authors or producers of a society as much as products of it (see Figure 2.5 below). Freire (1997, cited in Paiva & Braga, 2008) sees the role of the teacher being “not to transmit new knowledge, but to create new realms of possibility for students to produce and/or construct knowledge”.

![Figure 2.5 Products vs. Producers of society](image)

Chik (2007) discovered, through examination of learner narratives in Hong Kong, that a sense of identity seems to promote the development of autonomy and leads to a more satisfying learning experience overall. Likewise, long-term investigations by Huang (2009) of students in Hong Kong who metamorphose from second language learners to second language teachers, also reveal the close relationship between identity, agency and autonomy.
2.3.2 Reflection
Little (1999a, p. 27) believes that schooling should play a dual role in that it provides learners with knowledge, but at the same time enables them to critically question that knowledge; thus, in the language classroom, students should be able to use the FL as a tool for reflection. Little (1999b, p. 4) therefore believes that in order for autonomy to develop, there must be opportunities for reflection and self-assessment. He sees a shifting in power and control when reflection is encouraged, since he considers that “reflection entails learner initiative” and when there is “learner initiative”, there is the “beginning of learner control”. He also believes that reflection “necessarily entails self-assessment” (Little, 1999b, p. 6).

Dam & Little (1998, p. 128) view learning as a cyclical process of planning, implementation and evaluation, in which they see reflection as playing a “crucial role”. They see reflection as being at the centre, not just of language learning, but of education generally. By developing this capacity in the language classroom, we are enabling learners to acquire a skill which can be applied in “other domains of life”. It is a skill which can only be developed “gradually, on the basis of practice” (Dam & Little, 1998, p. 130).

Interestingly, Huttunen (2003, pp. 125-126) sees reflection as working on three different levels of:

(1) Mechanical reflection
(2) Pragmatic reflection
(3) Emancipatory reflection.
She has based these levels on the three “knowledge-constitutive interests” of Habermas (1972, cited in Huttunen, 2003), who, in turn, based his “interests” on Aristotle’s three dispositions of human behaviour. At the Mechanical level, learners simply take facts on board without attending to their relevance or linking them to past
knowledge and experience. At the Pragmatic level, the learner has an increased understanding of facts, due to an action or as a result of it, but is unable to analyse them further or relate them to experience. At the Emancipatory level, however, the learner gets a new perspective on things and gains fresh insights, while “engaging in reflection”. Connections are made with previous knowledge and experience and reasons are sought for actions and the outcomes of actions.

It seems evident that these three levels of reflection conveniently fit into the heteronomy-autonomy continuum, as shown in Figure 2.6 below.

<table>
<thead>
<tr>
<th>HETERONOMY</th>
<th>AUTONOMY</th>
</tr>
</thead>
<tbody>
<tr>
<td>MECHANICAL REFLECTION</td>
<td>PRAGMATIC REFLECTION</td>
</tr>
<tr>
<td>Nolen, 1995; Namenwirth, 1996; Waite-Stupiansky, 1997; Rujiketgumjorn, 2000</td>
<td></td>
</tr>
<tr>
<td>Huttunen, 2003</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.6** The heteronomy-autonomy continuum with Huttunen’s 3 levels of reflection

Such an approach to learning where emphasis is placed on the value of reflection, metacognitive awareness and assessment/evaluation, links with the ideas of Dam (Dam, 1990, p. 21; Dam, 1995, p. 31) where the questions her high-school students, in Copenhagen, Denmark, ask themselves in the evaluation stage are:

(i) What did I/we do?
(ii) What was good/bad? Why?
(iii) What can it be used for?

These questions also relate to the four questions which Little (1996a, p. 3; 1999a, p. 28) views as being essential to critical reflection (see Chapter 1, Section 1.3) and the fostering of autonomy.

Jiménez Raya (2006, p. 127) also believes that reflection is critical to the development of autonomy since ideas are “brought to consciousness”. This enables
the learner to consider an idea or problem from the perspective of an “outsider”, so that the learner becomes “his own critic” and can “identify weak spots”. This helps in the formation of goals and in leading the learning from what is known, to what is new. Once an idea or problem has been evaluated, it can be used as the basis on which to make informed choices about what to do or what not to do.

Sougari (1999, p. 157) asserts that teachers must encourage the processes of “critical reflection” in order to foster “independent and effective language learners”. Likewise, Cotterall (2000, p. 112) believes that “the potential for learner autonomy increases as an individual’s learning awareness grows. Therefore activities which prompt learners to reflect on their learning aim to enhance learners’ insights into their learning processes. This initiates autonomous learning”. Cotterall regards reflection as the “sine qua non” of autonomy in language learning since without it, learners cannot look back and assess the progress of their learning and neither can they modify their actions nor look forward and make plans for future learning activities.

2.3.3 Ownership
The question of ownership both of learning and of the foreign language would seem to be fundamental to the promotion of autonomy and yet it is rarely discussed in depth. It is perhaps a complex area to deal with because it brings in questions of learners’ acceptance of the foreign language, the processes of acculturation, as well as a number of affective factors.

Kohonen (2001b, p. 34) believes that very useful discussions can be formulated around the topic of “ownership” in relation to autonomy in language learning. It can be looked at from a political point of view in terms of the “power relationships between teachers and learners” or it can refer to the way that learners view their
language learning experiences as well as “the world around them”. Learners’ learning cannot be “meaningful” unless placed in a context where learners are able to “construct and interpret” it. Ownership, he thinks, also poses questions on an “emotional” level of who, in the learning situation, has the right or the responsibility to plan and take decisions. This would all depend on how much learners feel that they have responsibility to make choices, take control or take the initiative. Kohonen puts his ideas into diagrammatic form (adapted from Collis & Dalton, 1990, cited in Kohonen, 2001b, p. 35) as shown in Table 2.4 below, which shows ownership of learning as being on a spectrum or continuum of Teacher Ownership, Shared Ownership and Learner Ownership:

Table 2.4 Kohonen’s model of ownership on a continuum

<table>
<thead>
<tr>
<th>Teacher ownership</th>
<th>Shared ownership</th>
<th>Learner ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td>External control based on teacher authority.</td>
<td>Teacher encourages learner negotiation and helps them learn the appropriate skills.</td>
<td>Internal control by the learner, facilitated by the teacher.</td>
</tr>
<tr>
<td>Teachers see themselves as responsible for learning.</td>
<td>Shared responsibility through negotiation and learning contracts.</td>
<td>Learners see themselves as responsible for their own learning.</td>
</tr>
<tr>
<td>Encourages learner’s dependence on the teacher.</td>
<td>Interactive, negotiable dependence, with more learner choices.</td>
<td>Learner independence, and responsible interdependence.</td>
</tr>
</tbody>
</table>

From Kohonen (2001b)

Kohonen (2001b, p. 34) takes pains to stress that because greater ownership is assumed by learners, this does not necessarily mean that teachers then take a back seat in the learning process. Rather, in order for learners to be able to take greater
Banking of knowledge

Increased levels of understanding and application to professional practice to improve practice

Iteration between surface learning and deep learning

New knowledge construction through reflection

Surface learning

Deep learning

Banking of knowledge

Adapted from Black & Plowright (2008)

Figure 2.7 A learning continuum in which banking/ownership plays a key role
responsibility for their learning, there necessarily has to be “a careful balance between learner control and teacher support and feedback.” Crabbe (1999, p. 141) asserts that the key to improvement of individual performance, either of teaching or learning, is a greater sense of power and ownership, which will be “driven by an exploratory attitude and working within a curricular framework that is flexible and dynamic enough for individual explorations”.

Elsewhere (Black & Plowright, 2008), ‘ownership’ is referred to as the “banking of knowledge” and the role that ownership/banking plays in the transition from academic/surface learning to reflective/deep learning is made clear in Figure 2.7 above.

Black & Plowright (2008, p. 34) see banking/ownership of learning as being “an iterative process between deep and surface levels of learning” and like Kohonen’s model in Table 2.4 above, Figure 2.7 above shows the learning continuum and the difference in effect on learning, between traditional behaviourist approaches and non-traditional constructivist approaches, with a consequent increase in depth of learning.

Likewise, Reinders (2010, p. 52) sees the promotion of learner autonomy as being not just the developing of particular language skills, but rather requires “developing a certain mind-set that sees learning as an active process of discovery”. Reinders (ibid) is of the opinion that if learners realise that they are valued and supported, then they will be more likely to adopt this mind-set and teachers will regard learners’ ownership of the learning process as being important.

Dufeu (1994, p. 7) asserts that “language cannot be separated from its use”. Language users or “participants” can adopt the foreign language and “integrate it so
well that it becomes their own”. Because of this “direct contact” and ownership of the language, it no longer seems “foreign”.

Thus, moves towards autonomy help promote both ownership of learning, which becomes more meaningful, and, at the same time, promotes ownership of the language, which takes on personal significance.

2.3.4 Self-determination
Closely-linked to the concept of identity is that of self-determination. While identity has more to do with a way of being and acting in the world, through the means of language, self-determination is more related to a learner’s self-image or self-concept which is based on feedback received within the learning environment and the world at large and how (s)he decides to act based on that. Self-efficacy, on the other hand, has to do with the light that is shed on the learner’s participation in the language-learning process and the interpretation or understanding of her/himself that the learner makes of it. Williams and Burden (1997, p. 97) stipulate that the views of the world that individuals hold, have an influence on their self-concept, while, equally well, their self-concepts affect their way of viewing the world. It is the combination of these views which will determine the learner’s success in learning.

Also very closely related to self-efficacy and self-determination is the quality of self-esteem, with Deci and Ryan (1995, p.33) making a distinction between what they refer to as “contingent” self-esteem and “true” self-esteem. A person who constantly requires reassurance, as to their ability, has “contingent” self-esteem, while someone with “true” self-esteem is not so affected by the ups and downs of success and failure, “acting agentically” from their “integrated self”, while someone with “contingent”
Figure 2.8 A continuum of autonomy with increasing self-determination

Adapted from Ryan & Deci (2000)
self-esteem will be more affected by comparison with external and “socially imposed standards” (Deci & Ryan, 1995, p. 35).

Ryan and Deci link self-determination and motivated behaviour with autonomy and they view self-determination as being on a continuum of relative autonomy (Ryan & Deci, 2000, p. 72; Ryan & Deci, 2002, p. 16), (see Fig. 2.8 above).

Very interestingly, Deci & Ryan (1995, p. 33) make a remark about autonomy which is very similar indeed to that of Breen & Mann (1997, p. 141). While Deci & Ryan, speaking as educational psychologists, say that someone may feel forced to give up autonomy and a true sense of self, while taking on a socially implanted self, Breen & Mann, speaking as language researchers and educators, say that because learners wish to please their teachers, having discovered which behaviours their teachers wish to see, it is quite possible that they “will give up their autonomy to put on the mask of autonomous behaviour”. Schmenk (2006, p. 84), in a similar manner, argues that in some cases learners simply “mimic” autonomy, which is not true autonomy, but is, rather, a “façade” of autonomy.

Deci & Ryan (1995, p. 35) make the important distinction that “being autonomous is both an input to and a manifestation of the development of an integrated self and true self-esteem”. Thus, when Deci & Ryan (1995, p. 36) state that “only autonomous actions emanate from one’s true sense of self”, we again hear echoes of Breen & Mann (1997, p. 134) who say that the “autonomous person has a robust sense of self”.

It seems safe to conclude that Deci & Ryan’s continuum of self-determination can be added to the other categories in an expanded and inclusive version of our already-existing continuum of autonomy, as in Figure 2.9 below:
Having established identity, reflection, ownership and self-determination as factors of importance in the promotion and development of autonomy, however sporadically dealt with in the literature, it is now time to move on to investigating areas of consensus concerning autonomy among researchers in the field. Hurd (1998a, p. 219) asserts that it is essential to have agreement as to what we mean by autonomy in order to understand its effects on how we teach and how learners learn. She believes that a move towards autonomy has many significant repercussions. It seems advisable, therefore, before embarking on the AARP, to have a full understanding of what autonomy is exactly and what its promotion involves. This will be attempted by gathering together, from experts in the field, what seem to be verifiable areas of agreement.
2.4 Areas of consensus on autonomy

As will be seen in Section 2.5, many definitions and many different interpretations have been offered for autonomy in language learning, but if we are to arrive at a satisfactory definition of autonomy for the AARP, it would seem wise to look at what various experts have said about autonomy and try to create a catalogue of those aspects of autonomy on which there seems to be consensus. With the aims of clarity and succinctness, these areas of consensus will be arranged in a table, with the source and with the original citations, where these were given. In the last column, comments from the I-R and any further verifying sources will be offered, as appropriate. The first twenty-three entries in Table 2.5 are areas of common agreement taken from reliable sources, while the remaining seven are those identified by the I-R as also being of significance, with entry No.30, being of most particular relevance to the hypotheses of the present thesis.

Since the literature on ALL tends to be riddled with contentions, paradoxes, contradictions and “roiling inconsistencies” (Oxford, 2003, p. 75), establishing exactly where the particular areas of consensus lie, will make the job of finding a satisfactory definition of autonomy for the AARP somewhat simpler (see Table 2.5 below).
Table 2.5 Areas of consensus regarding autonomous language learning

<table>
<thead>
<tr>
<th>No.</th>
<th>Researcher</th>
<th>Area of consensus</th>
<th>Given citations*</th>
<th>I-R’s comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Benson (2006b)</td>
<td>...there are degrees of autonomy</td>
<td>Nunan, 1997</td>
<td>Benson, 2001; Broady &amp; Kenning, 1996; Candy, 1991; Gibbs, 1979; Grenfell &amp; Harris, 1999; Henner-Stanchina, 1976; Holec, 1981; Huang, 2009; Hurd, 1998a; 1998b; Huttunen, 1990; Karlsson et al, 1997; Kohonen, 1992a; Kumaravadivelu, 2003; Little, 1995a; McDevitt, 1997; Page, 1992; Paiva &amp; Braga, 2008; Sinclair, 1997; 2000 and Thanasopoulos, 2000 all concur that there are degrees of autonomy. Gardner &amp; Miller (1999) suggest there may be fluctuations in degree “over time and from one skill to another”. Farmer (1994), Farmer &amp; Sweeney (1994) and Pierson (1996) are also cited by Oxford (2008) as speaking of ‘degrees’ of autonomy, but she condemns such a conceptualisation as being too “simple” to explain such a complex phenomenon. This she does without offering a satisfactory alternative.</td>
</tr>
<tr>
<td>2.</td>
<td>Benson (2006b)</td>
<td>...the behaviour of autonomous learners ‘can take numerous different forms, depending on their age, how far they have progressed with their learning, what they perceive their immediate learning needs to be and so on’</td>
<td>Little, 1991</td>
<td>Kumaravadivelu, 2003</td>
</tr>
<tr>
<td>4.</td>
<td>Benson (2009)</td>
<td>...a multidimensional capacity that will take different forms for different individuals, and even for the same individual in different contexts or at different times</td>
<td>Benson, 2001</td>
<td>Huang, 2009. Some of these “dimensions” seem to be identifiable in Figure 2.9 above.</td>
</tr>
<tr>
<td>6.</td>
<td>Little (1991)</td>
<td>Autonomy is not a synonym for self-instruction; in other words, autonomy is not limited to learning without a teacher.</td>
<td>Little, 1990</td>
<td>Benson, 2001; Broady, 1996</td>
</tr>
</tbody>
</table>

*For Nos. 1 – 23, the ‘Given citations’ can be found at source (Researcher), given in the Bibliography, while for Nos. 24 – 30, the ‘Given citations’ can be found in the Bibliography.

Kumaravadivelu (2003) believes that the degree of autonomy appropriate for a particular teaching/learning context is something which can be negotiated between teachers and learners.
<table>
<thead>
<tr>
<th>No.</th>
<th>Researcher</th>
<th>Area of consensus</th>
<th>Given citations*</th>
<th>I-R’s comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.</td>
<td>Little (1991)</td>
<td>In the classroom context, autonomy does not entail an abdication of responsibility on the part of the teacher; it is not a matter of letting the learners get on with things as best they can.</td>
<td>Little, 1990</td>
<td>Dam, 1995</td>
</tr>
<tr>
<td>8.</td>
<td>Little (1991)</td>
<td>On the other hand, autonomy is not something that teachers do to learners; that is, it is not another teaching method.</td>
<td>Little, 1990</td>
<td>Dam, 1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Holec (1981) first defines autonomy as an &quot;ability&quot; and then he clarifies it further as a &quot;power&quot; or a &quot;capacity&quot;. Little (1991) and Aoki (1999) also refer to it as a &quot;capacity&quot;, as does the 'Bergen' definition, versions (a) Third Nordic Workshop (1990) and (b) Dam (1995).</td>
</tr>
<tr>
<td>13.</td>
<td>Sinclair (2000)</td>
<td>Complete autonomy is an idealistic goal.</td>
<td>Boud, 1981</td>
<td>Little, 1990; 1994; Nunan, 1997; Tschirhart &amp; Rigler, 2009. Sinclair (2000) sees the opposite of autonomy as a &quot;complete lack of autonomy&quot;; which she describes as &quot;vegetative&quot; (see Figure 2.3 above).</td>
</tr>
</tbody>
</table>

*For Nos. 1 – 23, the ‘Given citations’ can be found at source (Researcher), given in the Bibliography, while for Nos. 24 – 30, the ‘Given citations’ can be found in the Bibliography.

---

6 Little’s (1990; 1991) 4th negative defining principle of autonomy, which is “Autonomy is not a single, easily described behaviour.” is omitted since in No. 4 above, autonomy has already been described as “multidimensional”.

7 Holec (1985a) defines it as a “tendency”, “a dynamic process with a future” and a process of “autonomisation”, while Paiva & Brava (2008) see autonomy as “a dynamic system”.

8 In fact, Little reminds us that complete autonomy equates with “autism”.
<table>
<thead>
<tr>
<th>No.</th>
<th>Researcher</th>
<th>Area of consensus</th>
<th>Given citations*</th>
<th>I-R’s comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.</td>
<td>Sinclair (2000)</td>
<td>Autonomy can take place both inside and outside the classroom.</td>
<td>Little, 1996</td>
<td>This goes without saying.</td>
</tr>
<tr>
<td>21.</td>
<td>Karlsson et al. (1997)</td>
<td>Autonomy requires supportive structures, both internal and external.</td>
<td>No citations given.</td>
<td>This appears to be supported by Sinclair (2000) who says learners cannot just be put into situations where they need to be independent. The “internal” and “external” support could equate with Allwright’s (1990) “internal” and “external” resources, with “internal” coming from within the learner. Hunter and Cooke (2007) regard these “internal” resources as “self-knowledge”.</td>
</tr>
</tbody>
</table>

*For Nos. 1 – 23, the ‘Given citations’ can be found at source (Researcher), given in the Bibliography, while for Nos. 24 – 30, the ‘Given citations’ can be found in the Bibliography.

9 Holec refers to this process as “dynamic”.
10 Huang (2009) believes that the concept of autonomy can accommodate different interpretations and is universally appropriate, but this is disputed by Schmenk (2005).
11 It is important to distinguish the fact that autonomy is a “process” rather than a “product”.

42
<table>
<thead>
<tr>
<th>No.</th>
<th>Researcher</th>
<th>Area of consensus</th>
<th>Given citations*</th>
<th>I-R’s comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.</td>
<td>Everhard (2011)</td>
<td>Autonomy is generally considered to be something good, positive, valuable and desirable.</td>
<td>Allwright, 1990; Breen &amp; Mann, 1997; Kumaravadivelu, 2003; Everhard, 2006</td>
<td>Lewis (1978) and Lindley (1986) are philosophers rather than linguists, who believe that autonomy is essentially something good. A dissenting voice among educational philosophers is Hand (2006). Schmenk (2006; 2008) warns against its “idealization”. If autonomy is “good”, this would somehow imply that heteronomy is necessarily “bad”, while the two have a “dialectic” relationship.</td>
</tr>
<tr>
<td>26.</td>
<td>Everhard (2011)</td>
<td>Autonomy is not something that can be easily measured or evaluated.</td>
<td>Sinclair, 1999; Dam, 2001; Dixon, 2006; Mynard, 2006; O’Leary, 2007b</td>
<td>Lai (2001; 2008) has been studying this extensively. Cooker (2011), likewise, has been measuring autonomy in the context of SALL.</td>
</tr>
<tr>
<td>27.</td>
<td>Everhard (2011)</td>
<td>Autonomy is not something absolute.</td>
<td>Little, 1990; Nunan, 1997</td>
<td>This is a logical conclusion if we agree that there are degrees of autonomy. The same also applies to ‘heteronomy’, of course.</td>
</tr>
<tr>
<td>28.</td>
<td>Everhard (2011)</td>
<td>Autonomy takes time.</td>
<td>van Lier, 1996; Dam &amp; Little, 1998; Kohonen, 2003; Sambell et al. 2006</td>
<td>Housell (1979) states that the &quot;shift&quot; from other-directed to self-directed requires not just changes in &quot;methods&quot; but &quot;more fundamentally in his [the learner’s] perception of himself and his relationship to the world around him&quot;. Clearly, this will take time as &quot;this is less a change than a transformation of world-view&quot;.</td>
</tr>
<tr>
<td>29.</td>
<td>Everhard (2011)</td>
<td>Autonomy is challenging and difficult to achieve.</td>
<td>Little, 1990; Kumaravadivelu, 2003; Noels, 2007</td>
<td>Holec (1985a) asserts that a great deal of effort is required; care is necessary, but &quot;not to join in would be cowardly&quot;. Marshall (1996) says that it &quot;does not come easily to either teachers or learners&quot;, with Little (1990) asserting that it &quot;requires a lot of nerve&quot;.</td>
</tr>
<tr>
<td>30.</td>
<td>Everhard (2011)</td>
<td>Autonomy can be promoted through learner-centred assessment.</td>
<td>Holec, 1985b; Dickinson, 1987; 1992; Finch, 2003; Oscarson, 1989; 1997; Kohonen, 1999; 2000; Gardner, 2000; Little, 2007a; 2007b; Sambell et al., 2006</td>
<td>Harris &amp; Bell (1990) are educationalists rather than language teachers or linguists, but they see autonomy as increasing in degrees on a continuum (see Figure 2.12), as greater responsibility for assessment is taken on by learners, on a gradation from collaborative assessment, to peer-assessment, to self-assessment. Kostopoulou (2007; 2010) provides a very comprehensive overview of self-assessment, but chooses to ignore the value of peer-assessment, while Everhard (2011a; 2011b) makes strong arguments for both. Kato (2009) believes that monitoring and self-assessment in language learning are “essential elements” which raise awareness and promote autonomy. Little (2005c) believes self-assessment is “key” to the development of learner autonomy.</td>
</tr>
</tbody>
</table>

*For Nos. 1 – 23, the ‘Given citations’ can be found at source (Researcher), given in the Bibliography, while for Nos. 24 – 30, the ‘Given citations’ can be found in the Bibliography.
2.5 Defining autonomy

Trying to define autonomy is not an easy task (O’Brien & Guiney, 2001, p.54), but having established 30 areas of consensus from among the great diversity of opinions expressed by researchers on autonomy in EFL, in Section 2.4 (see Table 2.5), this should make the difficult task of defining autonomy just a little bit more straightforward. There are many reasons that contribute to making the task of defining autonomy difficult, not least of which is the multiplicity of terminology which is used to describe what seems to be essentially the same thing (Wright, 2005). Thus, we have terms like self-direction (Candy, 1991), learner-centredness (Tudor, 1996), individualization (McDonough & Shaw, 1993), self-regulation (Schunk & Zimmerman, 1994) and personal leadership (Farrar, 2007), to name but a few. Even when researchers do use the term ‘autonomy’, there is a tendency within the literature, which seems to stem from widespread insecurity, to constantly define and redefine what ‘autonomy’ actually is (Cotterall, 2008), out of an attempt to ensure that we are all talking about the same ‘version’ of autonomy.

Fenner (2000, p. 78) suggests that one of the reasons why autonomy is difficult to define is that it has more to do with an “attitude” or a philosophy, rather than any particular methodology. Paiva & Braga (2008, p. 444), on the other hand, suggest that it is the “complex” and “multifaceted” nature of autonomy that make it “virtually impossible to be comprehensibly described by a single definition”. In the next section, we will briefly examine some of the reasons why defining autonomy can be a challenge.
2.5.1 Multiple interpretations and definitions

In a way, all the confusion and controversy surrounding autonomy and all of its complexity simply add to the fascination and the challenge of being able to come to some understanding of it, with regard to both theory and practice (Everhard, 2006). Besides, Kojima (2006, p. 71) suggests that from a cultural point of view, it is useful to regard autonomy as being open to “different interpretations” and of therefore being “universally appropriate”. Very similar ideas are expressed by Breen & Mann (1997, p. 149), who suggest that multiple interpretations, rather than a consensus view of autonomy, are more conducive to innovation in teaching practices and also accommodate cultural diversity. However, this way of looking at autonomy seems to come into conflict with the position taken by Schmenk (2005, p. 111), who speaks in terms of autonomy perhaps not being a very appropriate concept to promote universally as it is in danger of being homogenized and of suffering from McDonald-like expansionism and of becoming globalised. Schmenk (2005, p. 115) states quite emphatically that “it is pivotal that people begin to reflect on the theoretical and practical background of autonomy as a cultural and political concept and seek to relocate it in specific social and cultural settings”.

Thus, Schmenk (ibid) believes that the only hope for autonomy lies in its becoming “glocalised”. This is an alternative view which seems to have been ignored as belonging to the ‘fringe’, but clearly it is a view which warrants further consideration and is of considerable relevance to the present research. Elsewhere, she warns against the emergence of what she refers to as “autonomy lite” as a result of trivialization and reductionism of the concept (Schmenk, 2008, p. 101, author’s emphasis), and its consequent “sloganization” due to its “idealization”. Schmenk
(2008, p. 111) argues that autonomy and heteronomy are “inextricably intertwined” in a “dialectic relationship” and she highlights for us the dangers of “illusions of autonomy” and “façading” (2006, p. 83). Clearly, in our exploration of autonomy and its promotion, we have to be aware of the possible traps. For this reason, we will follow the advice of Schmenk (2005) and aim to find a “glocalised” definition for the AARP. This we will do after examining six definitions offered by experts in the field, with a view to filtering and refining these in order to reach the definition most appropriate for the AARP.

**Holec’s definition**

Many researchers find it futile to try and improve upon Holec’s (1981, p. 3) original definition, where he refers to it as:

- … the ability to take charge of one’s learning.

Benson, one of the leading researchers in the field, till recently, had always been in favour of trying to find a satisfactory definition for autonomy since “we need to know what it is that we are trying to foster” (Benson, 2006, p. 736); however, more recently, in speaking of Holec’s definition, he says that while there seems to be general consensus that autonomy is indeed an “ability” or “capacity”, there is not so much consensus regarding what that capacity actually entails (Benson, 2009, p. 14). Interestingly, he concludes that “if autonomy can be defined at all, it must be defined as a *composite* of abilities, attitudes or dispositions” (Benson, 2009, p. 18).

Previously, in Section 2.3 of this Chapter, there was an attempt to examine some of the elements, such as reflection, a sense of identity, a sense of ownership and self-determination, which might constitute elements in such a “composite” of autonomy. This view, concerning abilities, attitudes or dispositions, echoes the view of Barnett
(2007, p. 29), already discussed in Chapter 1, Section 1.1, related to “authenticity” in a HE setting, and such a “composite” is of importance to the context of the AARP and also seems to clarify the need, in the present context, to expand on Holec’s original definition.

**Little’s definitions**
Apart from Holec’s, now legendary, definition, Little’s (1990, p. 7) ‘negative’ definition, which describes for us what ALL is not, rather than what it is, has been the most popular definition, according to Benson (2006a, p. 737). The five points Little offers us, by way of definition, are:

- Autonomy is **not** a synonym for self-instruction; in other words, autonomy is **not** limited to learning without a teacher.
- In the classroom context, autonomy does **not** entail an abdication of responsibility on the part of the teacher; it is **not** a matter of letting the learners get on with things as best they can.
- On the other hand, autonomy is **not** something that teacher’s do to learners; that is, it is **not** another teaching method.
- Autonomy is **not** a single, easily described behaviour.
- Autonomy is **not** a steady state achieved by learners (author’s emphasis).

Little (1991, p. 4) also offers what he refers to as a “‘provisional’” definition, which he chooses never to refine or finalise, as follows:

- Essentially, autonomy is a capacity for detachment, critical reflection, decision-making, and independent action. It presupposes but also entails, that the learner will develop a particular kind of psychological relation to the process and content of his learning. The capacity for autonomy will be displayed both in the way the learner learns and in the way he or she transfers what has been learned to wider contexts.
Little, according to Carson (2010, p. 81), sees the “key” to the promotion of autonomy as being learner involvement, learner reflection and target language use. These are elements which were all of great importance in the AARP, as are Little’s four levels of understanding, necessary for the development of autonomy, which are outlined in Figure 2.11, in Section 2.6.2, and are essential for the AARP model. Also, of particular importance, in the way that peer- and self-assessment processes were conducted in the AARP, is the “capacity for detachment” which Little refers to in his definition above, since, unlike other assessment studies, feedback and conferencing were avoided before and during rating, in order to encourage such “detachment”.

*Breen and Mann’s definition*

Breen and Mann (1997, p. 134) begin by defining the characteristics of the autonomous learner and then they give their definition of autonomy, where they make clear the difference in their stance from that of Holec’s:

- Eight different qualities we see as characterizing being autonomous: the person’s stance towards the world, their desire for what it is they are learning, their sense of self, their metacognitive capacity, their management of change, their independence from educational processes, their strategic engagement with learning, and their capacity to negotiate.
- … autonomy is seen as a *way of being* in the world; a *position from which to engage* with the world. Unlike Holec, we are proposing that autonomy is not an ability that has to be learnt, but a way of being that has to be discovered or rediscovered.

Thus, Breen and Mann offer us quite a clear and comprehensive profile of the autonomous language learner and quite a unique view of autonomous language learning.
Macaro’s definition

The definition which Macaro (1997, p. 168) offers us is based on both pilot and extended studies of foreign language learners, in Italy, with learners of EFL and in English schools with learners of MFL, which he has undertaken. He defines autonomy as:

- … an ability which is learnt through knowing how to make decisions about the self as well as being allowed to make those decisions. It is an ability to take charge of one’s own language learning and an ability to recognise the value of taking responsibility for one’s objectives, content, progress, method and techniques of learning. It is also an ability to be responsible for the pace and rhythm of learning and the evaluation of the learning process.

Macaro, like Little (1990, p. 7; 1991, p. 4) and van Lier (1996, p. 13), places emphasis on decision-making, which is very relevant to the assessment procedures which formed such a significant part of the AARP.

Allwright’s definition

Allwright (1990, p. 1) defines autonomy in some detail, as follows:

- Ideally, autonomy is a state of maximal self-development in which the individual has developed his or her own inner resources to the full, and is therefore as self-sufficient as it is personally appropriate to be, but where the individual also:
  (a) recognises needs that can only be met by recourse to external resources;
  (b) can identify such needs as they arise;
  (c) knows how to gain access to the appropriate external resources;
  (d) knows how to make optimal use of such external resources;
  (e) and can do the above without limiting the autonomy of others, and without unduly compromising his or her own autonomy.
• Autonomy is therefore a state of optimal equilibrium between dependence and self-sufficiency.

Allwright’s definition is interesting, because he regards autonomy as a process of development in which the individual makes maximum use of both internal and external resources, something which the I-R was aiming to promote on the AARP, by raising awareness, through analysis of modality strengths (see Appendix 1), of left/right brain orientation (see Appendix 2), and of individual strengths and weaknesses, through Learner/Teacher Contracts (see Appendix 4).

Paiva’s definition

Paiva’s definition takes into account complexity and chaos theories and therefore covers many elements which are of importance in developing autonomy through the AARP. Paiva (2006, cited in Paiva & Braga, 2008, p. 449) defines autonomy as follows:

• … a complex socio-cognitive system, subject to internal and external constraints, which manifests itself in different degrees of independence and control of one’s own learning process. It involves capacities, abilities, attitudes, willingness, decision-making, choices, planning, actions, and assessment either as a language learner or as a communicator inside or outside the classroom. As a complex system it is dynamic, chaotic, unpredictable, non-linear, adaptative, open, self-organizing, and sensitive to initial conditions and feedback.

Paiva is to be admired in that he has attempted to create an all-inclusive, comprehensive definition. Many of the elements included in his definition are in harmony with those the I-R has selected for the AARP definition, which is offered in Table 2.6, below, and which takes into account the conditions in which the research
was conducted, the previous knowledge and experience of the participants, the aims of the Language Mastery I course, and the context of studying in the SOE, AUTh.

2.5.2 The AARP definition of autonomy

In creating the definition for the AARP, the I-R has tried to accommodate both the universal features of autonomy and the autonomous learner which come out of the six definitions above, but at the same time to recognise the need to pay heed to the warnings of Schmenk (2005), outlined previously in Section 2.5.1, who fears the “globalisation” and “MacDonaldisation” of autonomy. Thus, in the case of the AARP, the global and the local have been combined to create a “glocal” definition to suit a local ‘version’ of autonomy. The AARP definition, which draws mainly on Allwright (1990), Breen & Mann (1997) and Macaro (1997), is presented in Table 2.6 below:

<table>
<thead>
<tr>
<th>Table 2.6   The AARP definition of autonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy is a way of being and a sense of self which is achieved through acquiring the ability to cooperatively make decisions about one’s own learning and that of others and which is exercised by being allowed to make and execute those decisions through access to both internal and external resources. The degree of autonomy achieved and exercised varies according to the disposition and predisposition of the learners in terms of affect, motivation, commitment, engagement, interaction, cooperativeness, ownership, reflection and uptake, and fluctuates according to circumstances.</td>
</tr>
</tbody>
</table>

2.5.3 The answer to the 1st preliminary research question

With the offering of the definition for the AARP in Table 2.6, the answer to the first of the two preliminary research questions outlined in Chapter 1, Section 1.3, seems to have been found, since the definition is at one and the same time sufficiently universal in its ethos, and, at the same time, appropriately local in its applicability.
The aim in the sections which follow will now be to resolve the second preliminary research question, outlined in the same section mentioned above.

2.6 Finding a model for the AARP

For the purposes of this research study, there has been an attempt to assemble the ideas of various scholars in the field and to assimilate where the areas of agreement lie between what can sometimes appear to be diverse, conflicting or contradictory views of autonomy. It becomes clear that there is a much greater amount of agreement amongst theoreticians and practitioners of autonomy than might at first have been apparent.

In the course of this assimilation of ideas from various experts, by using a heteronomy-autonomy continuum as the basis for a model of autonomy, there has been an attempt to contrive two things simultaneously:

(1) to demonstrate to what extent and in what way the various theories of autonomy overlap and fit into the heteronomy-autonomy continuum;

(2) to indicate how some of the many elements or ‘facets’ of autonomy, such as reflection, ownership and self-determination, which have been presented by experts as being on a cline which varies in degree, can also be conveniently slotted into the continuum model.

The main aims of the chapter, so far, have therefore been analysis and synthesis. This will culminate, in the sections that follow, with the finding of a more comprehensive, yet flexible model for the AARP. The kind of model we will be looking for is one which shows the variables which interplay and interact in the learners’ “social and physical worlds” (Noels, 2009, p. 296) and those elements, with regard to autonomy, which “nurture or thwart its unfolding” (Noels, 2009, p. 302).
In the section which follows we will take a look at some of the models which have been suggested for ALL and consider their suitability for the AARP.

### 2.6.1 Models of autonomy

Given the amount of enigmaticity and elusiveness (Kohonen, 2001b; Everhard, 2006; Little, 2010) attached to autonomy in language learning, as a linguistic theory it is very difficult to conceptualise and it is even more difficult to commit these conceptions to some kind of diagrammatic form on the computer screen or on paper. Nevertheless, it is useful to have a diagrammatic representation of autonomy and in the sections which follow, four such diagrammatic models will be considered, namely those of Nunan (1997, p. 195; 2000, pp. 11-12), van Lier (1990, p. 41), Kohonen (2001b, p. 50) and Littlewood (1997, p. 83).

In his review article of autonomy, Benson (2006b) attempts to make some comparisons between models, though not all of those he compares were presented by their creators in diagrammatic form. Benson’s main interest lies in the levels or phases of autonomy represented in the models of Nunan (1997), Littlewood (1997), Macaro (1997), Scharle & Szabó (2000) and Benson (2001). The first two models mentioned by him are of particular interest to us because of what Benson (2006b, p. 24) refers to as the “possible progression from ‘lower’ to ‘higher’ levels of autonomy which they imply”. Nunan’s (1997, p. 195; 2000, pp. 11-12) model becomes more comprehensible to us if we begin by looking at Nunan (2000, p. 4), where he describes his ongoing work with Benson and their investigation of language learning “careers”. This work helps shed light on Nunan’s model, which will be reproduced in an adapted form in Table 2.7 below.
**Nunan’s model of autonomy**

The usefulness of Nunan’s (1997; 2000) model for autonomy will become more transparent if we take into account the fact that Nunan (2000, p. 11), like the I-R, views autonomy in language learning as being:

1. a matter of degree
2. on a continuum
3. something which can be developed in learners.

Another significant factor is that, like Falchikov (2007), Nunan (2000, p. 5) views progress with learning as something akin to an apprenticeship, with the language learner progressing from “apprentice”, to a “competent” practitioner and eventually to a “master” practitioner. Nunan (2000, p. 8) believes that there is a “fascinating, but puzzling relationship” between autonomy and communication, and that autonomy grows together with an increasing communicative conceptualisation of language. Nunan (2000, p. 8) sees EFL learners as moving through four developmental stages, where the learner considers:

1. English is a subject much the same as any other in the curriculum
2. English is important
3. English is different from other subjects in the curriculum: it is a tool for communication that can be used outside the classroom
4. In order to develop my ability to use English for communication, I need to take greater control of my own learning.

Such control is concerned with management, cognition and also content (Nunan, 2000, p. 8). In the model provided in Table 2.7 below, adapted from Nunan (1997, p. 195; 2000, p. 11), we can see the progression or gradation towards autonomy as being on a continuum from Levels 1 to 5, rather similar to the four developmental stages listed above, with levels described in terms of Learner Action, the Content of
Learning and the Process of Learning. While this model is a useful one in terms of illustrating autonomy as evolving through stages, at the same time it could give the neophyte instructor the impression that autonomy evolves through clearly-defined stages, in lockstep with the progression of the learners in their language-learning, something which Kumaravadivelu (2003, p. 144) states is fallacious. Oxford (2008, p. 47) is also critical of Nunan’s model as she feels that what she sees as his “metaphor” for describing autonomy, using the idea of “stages”, does not “fully describe the complexities of a given learning situation”; nevertheless, she seems to reconsider this and decides to modify his model herself (Oxford, 2008, p. 45-46), adding a non-stage, and extending the categories. Although Oxford (2008, p. 44) may feel that she has improved upon Nunan’s “visual metaphor”, for our present purposes, it is more convenient to retain Nunan’s original entries, but turn them around to fit the autonomy-heteronomy continuum (see Table 2.7 below).

**Table 2.7 Nunan’s model of autonomy**

<table>
<thead>
<tr>
<th>Level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner action</td>
<td>Awareness</td>
<td>Involvement</td>
<td>Intervention</td>
<td>Creation</td>
<td>Transcendence</td>
</tr>
<tr>
<td>Content</td>
<td>Learners are made aware of the pedagogical goals and content of the materials they are using.</td>
<td>Learners are involved in selecting their own goals from a range of alternatives on offer.</td>
<td>Learners are involved in modifying and adapting the goals and content of the learning programme.</td>
<td>Learners create their own goals and objectives.</td>
<td>Learners go beyond the classroom and make links between the content of classroom learning and the world beyond.</td>
</tr>
<tr>
<td>Process</td>
<td>Learners identify strategy implications of pedagogical tasks and identify their own preferred learning styles/strategies.</td>
<td>Learners make choices among a range of options.</td>
<td>Learners modify/adapt tasks.</td>
<td>Learners create their own tasks.</td>
<td>Learners become teachers and researchers.</td>
</tr>
</tbody>
</table>

*Adapted from Nunan (1997)*
The consensus view in the literature is that autonomy is something that fluctuates from learner to learner, from situation to situation, from task to task and even from one minute to the next. Although it provides a useful continuum, and the changes in the learners’ mind-set becomes clearer in the adaptation in Table 2.7, Nunan’s model, unfortunately, does not seem to allow for the notion of fluctuation, but, rather, seems to progress through clearly-defined stages, to which Oxford (2003), as previously mentioned, objects, so would not be suitable for the AARP. Another way in which this particular model is limited is the fact that we cannot see the role of the instructor in the process as it only focuses on the role of the learner. This same limitation applies to the models presented in Figure 2.10 below, with the exception of Kohonen, who includes the teacher and the institution.

A comparison of three more models
The three models of van Lier, Kohonen and Littlewood, displayed in Figure 2.10, show the elements which these researchers recognize as interweaving to encourage or foster autonomy. In van Lier’s (1990) model, we see a very vibrant interplay, on pre-cognitive and post-cognitive levels, between Awareness (which also features in Nunan’s model, see Table 2.7), Autonomy and Authenticity, the latter being an element which is also highlighted by Newby (2000, p. 20) as being of importance in promoting autonomy. The three As are all pivoted on social interaction, with engagement and intake coming into play when the three As overlap.

In Littlewood’s (1996, p. 432; 1997, p. 83) model, we see Motivation, Confidence, Knowledge and Skills lying at the heart of a circle or cycle where there is constant interplay between the three encapsulated levels of autonomy as a Learner, a Communicator and a Person. Although this model is an interesting one, it does not
van Lier’s view of autonomous language learning and its interrelationship with awareness and authenticity.

Littlewood’s view of the three types of autonomy as a person, learner and communicator and their interrelationship.

Kohonen’s view of experiential learning and its link with autonomy.

Figure 2.10 Three models for autonomous learning offered by van Lier (1990), Littlewood (1996; 1997) and Kohonen (2001b)
project the possibility of progression or regression in degrees, and of fluctuation, which is so necessary for the AARP.

Like Nunan and van Lier, Kohonen’s (2001b, p. 50) model places Awareness at the heart of autonomous or experiential learning and he sees Awareness as functioning on the three levels of Personal, Process and Task, which form the three points of a triangle. He also highlights the importance of the Awareness of the Instructor and of the cultural climate both within the institution and society. Again, the model is an interesting one, but has moved away from his previous tabular representations where there was progressive responsibility and ownership of learning (see Table 2.2 and Figure 2.7) and there is likewise no feeling of progression from other-directed to self-directed learning (see Figure 2.3). Unfortunately, apart from Nunan’s (1997, p. 195; 2000, p. 11) model (see Table 2.7), none of the four models discussed, convey the notion of degrees of autonomy or demonstrate a clear continuum along which the various elements intertwine to create the variability or fluctuating degrees of autonomy which are of interest to us in the present project. Since no satisfactory model has been found within the EFL literature, which meets the criteria necessary for the AARP, it will be necessary to look beyond the borders of the language learning literature and see if any existing models within other sciences could be modified or adapted.

2.6.2 The proposed AARP model
A key element which is a requirement for the model is that it should be able to accommodate what Little (1996a, p. 3; 1999a, p. 28) states are the four levels of necessary understanding on the part of learners if they are to accept and sustain responsibility on their path to autonomy (as shown in Figure 2.11 below). The four questions on which the levels are based are given in Chapter 1, Section 1.3 and the same four questions/levels are included in the final AARP model (see Table 2.8 below).
These four levels, and the last in particular, are of utmost importance to the AARP. For this reason, we will look beyond the field of FL and EFL education and we will modify and adapt, with permission, a framework created by Stolk, Martello and Geddes (2007), which was designed to serve their interest in the lifelong learning of engineers. This modified framework seems to provide a satisfactory model for the AARP, as well as what Oxford (2008, p. 44) refers to as a “visual metaphor”, which can be seen in Table 2.8 below.

The usefulness of this model, though it originated from a non-EFL context, lies in the fact that it views autonomy as being developed or increased by degrees on a continuum, but at the same time nothing is fixed or static and the model allows for regression as well as progression, and also permits of considerable fluctuation. The four plains on which it functions coincide, symptomatically, with the four levels of understanding prescribed by Little (1996a, p. 3; 1999a, p. 28) and it therefore seems to meet the needs of the AARP perfectly.

This model or framework also accommodates all the theories of autonomy which think in terms of a continuum or gradation, such as Smith’s (2002) weak-strong version, or Littlewood’s (1996; 1997) reactive-proactive learner, together with the conceptualisations of Holec (1981),
Kohonen (1992a; 1992b), Huttunen (1986; 1990) and Sinclair (2000) (see Figures 2.3 and 2.4).

This is because the autonomy continuum progresses from None, to Low, to Medium, to High (see Table 2.8 below).

Most important of all for the present study is the fact that this model also matches perfectly with that of Harris & Bell (1990, p. 111), who talk about and diagrammatically show a continuum of assessment which gradates from teacher-controlled assessment at the one end to learner-controlled assessment at the other (see Figure 2.12 below). Harris & Bell refer to “degrees of learner autonomy in assessing”, which is reminiscent of the term “assessment autonomy” with its “fading of support”, used by Falchikov (2007, p. 139), and they show a progression from traditional assessment to collaborative assessment, from collaborative assessment to peer-

![Figure 2.12](image)

*Adapted from Harris & Bell (1990)*

**Figure 2.12** A continuum of assessment offering degrees of autonomy

assessment, and from peer-assessment to self-assessment, with the latter two being regarded as very closely-related (see Figure 2.12 above) and, coincidentally, matching the progression in the AARP.

The AARP model, given in Table 2.8 below, also accommodates the various autonomy-heteronomy continua as well as the various other models of learning and autonomy which we looked at earlier in the chapter, with regard to teacher/learner control, progression in ownership of
learning, as well as progression in reflection (see Figures 2.1 - 2.9, 2.11 & 2.12, as well as Tables 2.2 - 2.4 & 2.7). None of this is imposed on the learners and, indeed, they can fluctuate, progress, or, indeed, regress in these spheres depending on their disposition, the task in hand and their willingness to assume responsibility.

As Stolk et al. (2007) state, the rows in the model represent areas of responsibility for learning, while the columns indicate varying levels of control in these areas. The beauty of this particular model for the AARP is that it shows how although each student was given equal opportunities for learning and engagement with learning, it is likely that there was considerable variation in the degree of autonomy achieved on the various levels by different individuals, on the various assignments and activities, but that the potential was there for all to reach the highest degree on each level (see Table 2.8), or to achieve what Reeve, Bolt and Cai (1999, p. 540) refer to as “rich autonomy”, combined with what Falchikov (2007, p. 139), previously mentioned, refers to as “assessment autonomy”. It also shows the clear progression Nunan (2000, p. 5) had in mind for EFL learners, moving from “apprentice” to “competent practitioner” to “master practitioner”. A comparison of the proposed AARP model with Oxford’s (2008, pp. 45-46) makeover of Nunan’s (1997, p. 195; 2000, p. 11) model (see Table 2.7 above) gives the impression that the AARP model might be what Oxford was aiming for, but her model, based more on theory than practice, lacks the essential element of assessment.
<table>
<thead>
<tr>
<th>Areas of Responsibility for Language Learning</th>
<th>Intellectual heteronomy</th>
<th>Academic Autonomy</th>
<th>Academic Autonomy</th>
<th>Intellectual autonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NO AUTONOMY</td>
<td>LOW DEGREE OF AUTONOMY</td>
<td>MEDIUM DEGREE OF AUTONOMY</td>
<td>HIGH DEGREE OF AUTONOMY</td>
</tr>
<tr>
<td>WHAT? Content, knowledge, skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Language instructor designs the syllabus and designates the list of language items and skills to be learned</td>
<td>Language instructor identifies language learning needs and provides problem statements</td>
<td>Collaborative identification of language learning needs or problems</td>
<td>Learners identify learning problems/needs and organize necessary knowledge or skills</td>
<td></td>
</tr>
<tr>
<td>• Language Instructor or institution or higher authority selects course materials</td>
<td>Language instructor defines learning content, skills and language acquisition goals</td>
<td>Language acquisition goals are flexible and collaboratively planned and modified</td>
<td>Learners establish language content acquisition goals based on defined needs and objectives</td>
<td></td>
</tr>
<tr>
<td>• Language instructor determines pace, mode and style of instruction</td>
<td>Low levels of ambiguity or problem-solving in learning content</td>
<td>Language instructor defines broad constraints and uses exploratory/problem-solving approaches</td>
<td>Learners select appropriate materials and comfortable pace of learning</td>
<td></td>
</tr>
<tr>
<td>WHY? Motivation, context</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Amotivation or extrinsic motivation</td>
<td>Extrinsic motivation with some student regulation based on perceived task value</td>
<td>Balance of extrinsic and intrinsic motivation</td>
<td>Intrinsic motivation derived from curiosity, passion, interest</td>
<td></td>
</tr>
<tr>
<td>• Emphasis on rote-learning and memorisation of facts, with little experimentation or practical application</td>
<td>Framework for language learning context provided by instructor, with connection points</td>
<td>Instructor in touch with student needs, background and interests</td>
<td>Learners relate activities to broader social and cultural contexts and values, making links with the real world</td>
<td></td>
</tr>
<tr>
<td>• Learning context as defined by the language teacher or syllabus</td>
<td>• Language instructor designs the syllabus and designates the list of language items and skills to be learned</td>
<td>• Learners establish language learning needs or problems</td>
<td>• Learners accept responsibility for actions and judgments</td>
<td></td>
</tr>
<tr>
<td>HOW? Strategies, process</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• The learning pathway is clearly defined by the language instructor</td>
<td>Language instructor suggests multiple strategies and approaches.</td>
<td>Language instructor as a resource, facilitator, enabler, and guide</td>
<td>Instructor as counsellor</td>
<td></td>
</tr>
<tr>
<td>• Emphasis is on reproduction of language instructor’s methods</td>
<td>Language instructor provides the learning resources, the learning plan and the scaffolding</td>
<td>Language instructor encourages multiple solutions and guides construction of meaning and understanding as required</td>
<td>Learners choose own learning pathway</td>
<td></td>
</tr>
<tr>
<td>• Learners follow instructor’s directions</td>
<td>Language instructor assigns tasks, readings and goals, offering minimum choice and selection</td>
<td>Negotiation with other learners and instructor, using inner and outer resources</td>
<td>Learners identify and evaluate learning resources</td>
<td></td>
</tr>
<tr>
<td>• Language instructor provides learning approach and solutions</td>
<td>Minimal development of alternative solutions, schemes of work and learner ownership and responsibility</td>
<td>Shared ownership and responsibility</td>
<td>Learners select and implement strategies to make use of resources</td>
<td></td>
</tr>
<tr>
<td>• Ownership lies with instructor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WITH WHAT SUCCESS? Feedback, evaluation, assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Language instructor evaluation</td>
<td>Language instructor defines assessment mechanisms</td>
<td>Collaborative assessment, involving peer, self and instructor</td>
<td>Ongoing realistic self-assessment of learning achievements and success</td>
<td></td>
</tr>
<tr>
<td>• Criteria for assessment may be hidden</td>
<td>Learners provide some justifications for answers and solutions</td>
<td>Internalisation of feedback</td>
<td>Internal incentives and rewards</td>
<td></td>
</tr>
<tr>
<td>• Use of M/C and right/wrong answers</td>
<td>Primarily extrinsic rewards or punishments</td>
<td>Understanding of abilities through pragmatic reflection, cooperation, banking of knowledge and higher order thinking.</td>
<td>Thorough reflection on decisions, learning outcomes and performance</td>
<td></td>
</tr>
<tr>
<td>• No learner self-reflection</td>
<td>Minimal, mechanical self-reflection</td>
<td>• Learners accept responsibility for actions and judgments</td>
<td>Learners’ ownership of learning and language</td>
<td></td>
</tr>
</tbody>
</table>

Based on Stolk, Martello & Geddes (2007) and Little (1996a; 1999a)
2.6.3 The answer to the 2nd preliminary research question

It would seem that the 2nd preliminary research question which was posed in the introduction to this chapter, in Section 2.0, has been resolved by the model found for the AARP. Provided we do not regard the gradations in autonomy as being rigid or fixed points on a scale, and thus allowing for the natural dialectics between autonomy and heteronomy, on which Schmenk (2006) insists, we can see that it can accommodate all learners and most learning situations where some kind of ongoing evaluation or assessment, evolving from teacher-directed to more learner-directed, is in force.

2.7 Concluding remarks

In this chapter, various theories and conceptualisations related to ALL were examined for their similarities and for their suitability to fit a simple model based on an autonomy-heteronomy continuum. A thorough search of the autonomy literature helped establish the main areas of consensus in relation to autonomy and thirty areas of consensus were identified and listed. Six definitions of autonomy from the EFL literature were highlighted and a suitable definition for the AARP was suggested, in answer to the 1st preliminary research question. Finally, four diagrammatic models of autonomy in EFL, were examined for their suitability for adaptation to the AARP, but all were rejected in favour of one from a non-EFL context which offered more flexibility, within a continuum of autonomy-heteronomy, capable of accommodating learners of all types and dispositions by allowing for fluctuation, progression and even regression. In this way, the 2nd preliminary research question was resolved.

The next chapter offers an overview of the assessment literature and attempts to reveal the crucial link between assessment and autonomy.
Chapter 3
Assessment for autonomy: an overview

3.0 Introduction
While the aim in the previous chapter was to set the background for the research study from the perspective of autonomy, the aim in the present chapter will be to set the background for the AARP from the perspective of assessment. More specifically, in the first section, there will be an attempt to describe the learning landscape from which the participants in the AARP, which was conducted in the School of English (SOE), AUTh, emerged. This will hopefully be achieved by a description of the influence which Testing Evaluation and Assessment (TEA) has on learners, teachers and society in general, as well as the beliefs which are specific to the EFL learners in the Greek HE context of the SOE, AUTh. The far-reaching influence of the Council of Europe (COE) and its production of the European Language Portfolio (ELP) and the Common European Framework of Reference (CEFR) will also be considered as will some of the key figures in the formulating of ideas and of materials which aim to promote autonomy through evaluation and assessment. The importance of peer-assessment and self-assessment, which feature significantly in the AARP, as a means to promote autonomy will be considered, with reference to the work of two of their main promulgators worldwide. The chapter will end with an overview of the key studies, which have investigated the use of peer- and self-assessment in writing skills and speaking skills, and their bearing on the AARP will be highlighted.

3.1 Testing, evaluation and assessment (TEA)
The aim here is to formulate a picture of traditional approaches to TEA in Greece and elsewhere and to outline the circumstances that have led to the introduction of alternative forms of assessment elsewhere, particularly in the language education
sector. Let it be clear that in these discussions of autonomy and of assessment, the aim is not to criticise any particular educational system or any particular approaches to testing, since faults and inadequacies can always be found in any system within any country, but rather the aim is to describe the learning environment, and, in particular, the EFL learning landscape from which the subjects used in this research study have emerged, as they enter courses of instruction with the instructor-researcher (I-R), and to describe the effects which a change of approach towards testing, evaluation and assessment (TEA) in the AARP, albeit short-term, may have had upon them.

3.1.1 TEA and its influence on learning

TEA, like autonomy, could be described as a “slippery” (Tschirhart & Rigler, 2009, p. 71) subject to deal with, since, as Kvale (2007, p. 57) asserts, it is “a field of contradictions”. Allwright & Hanks (2009, p. 19) claim that assessment is a “huge” topic and make the very insightful comment that the way in which language instructors view their learners has a great deal to do with the way in which they assess them. Likewise, Freeman (1995) and Boud & Falchikov (2006) believe that the way that learners behave and interact with the learning process has a great deal to do with the system of assessment deployed. This viewpoint is reiterated time and again in the literature related to study and learning in institutions of higher education worldwide.

Earl (2003, p. 13) asserts that there is a likeness between assessment, teaching and learning in that assessment is not a “singular entity”, but, rather, is “complex and dynamic” and she feels that assessment is worthy of our attention and interest and “deserves to be differentiated and understood in all of its intricacy”. Indeed, the interrelationship between teaching, learning and assessment in language learning is a very interesting one which has been the subject of debate for decades.
In the same way that entrenched ideas and practices in learning have made it difficult to introduce change, through autonomy, so it has been with TEA. In fact, attempts in the past to question, reform or revolutionise and to promote assessment and evaluation as “a part of learning” (Earl, 2003) such as is suggested in the work of Bloom, Hastings and Madaus (1971, cited in Earl, 2003), resulted in those voices calling for change and reform of the system being marginalised or silenced “by the power of testing as a mechanism for social control and social mobility” (Earl, 2003, p. 8).

A system of testing and examining where control is in the hands of a “higher authority” is so entrenched, not only amongst the pedagogic and academic community, but within society as a whole, that there is very great reluctance to question that authority and very strong resistance to shifting the balance of that power. Lewkowicz & Moon (1985, p. 62) remind us that “the right to formally evaluate others is a powerful weapon of control”, while Gibbs (1999, p. 41) believes that it is “the most powerful lever teachers have to influence the way students respond to courses and behave as learners”.

Strong resistance to change in assessment is highlighted metaphorically by Broadfoot (2001, cited in Earl, 2003, p. 11), who states that “we cling to the familiar like a much-loved garment, even when it is long past its best and ought to have been discarded long ago”. A similar notion concerning abhorrence of change is mentioned by Davies (1993, p. 125) when he states that “humankind seems naturally resistant to change. No matter how obvious it is, we cling to outmoded concepts, outmoded principles, outmoded values and continue to maintain our own self-interest”.
Earl (2003, p. 11) herself acknowledges the degree of power and control which examinations, and we can therefore assume, the practices of examination boards, have on the personal and collective consciousness, remarking that:

existing assessment practices are so deeply rooted in our collective intellectual and political consciousness that they have been almost impossible to challenge. They may even become more entrenched as policy makers realize the power of large-scale assessment to monitor, compare and leverage educational systems.

In recent years there has been a lobby of educators who have been interested in making the move from ‘testing of learning’ to ‘testing for learning’, and Paris (1998, p. 193) highlights for us the reasons why this has come to be:

Perhaps the most fundamental problem with standardised achievement tests is that they distort students’ motivation and learning by overemphasizing the importance of the scores as outcomes and measures of students’ abilities. Tests can redefine students’ goals for learning in counterproductive ways that make the outcome more important than learning as enquiry, reflection, and process.

The far-reaching effects of assessment are stressed by Gibbs (2006, p. 23), who reminds us that “assessment frames learning, creates learning activity and orients all aspects of learning behaviour”. Although, in the case of the AARP, students’ attitudes towards and beliefs about assessment may be adversely influenced by their previous experiences of it, optimistically, Ecclestone & Pryor (2003, p. 484) assert that although instructors cannot change these “prior experiences”, they can, however, help students to “re-evaluate these experiences”. In the sections which follow, exactly what these student beliefs are and how they are formulated, will be examined.

3.1.2 TEA and its influence on language learning

Davies (1990, pp. 1-2) makes the point that language testing enjoys the privilege of being “rightly central” to language teaching and also that it belongs within the
discipline of Applied Linguistics, but he is concerned about the interrelationship of the three and of “just how far testing is servant and how far master”. This same interrelationship appears to trouble other experts in the field, such as Baker (1989, p. 2) who talks about the shifts of emphasis which have been taking place in language teaching and the consequences that these had for language testing, but how the latter is somehow “more resistant to change” than is the case with teaching methodology and course design. Rust (2002a, p. 146) notes also that while there have been significant changes in teaching methods in higher education, this has not been matched by changes in assessment, due to “a significant lag”, which is a pity when there is such “strong evidence that involving students in the assessment process has definite educational benefits” (Rust, 2002b, p. 4).

This delay or ‘resistance’ to change is also confirmed by Little, who remarks that changes to public examinations in view of Communicative Language Teaching (CLT), for example, were “relatively slight” (Little, 2003, p. 225). The effects which TEA have on learners are often mentioned in the literature, particularly in terms of the fear and negative affect which they generate (Berry, 2008).

Boukouvalas (2001), in writing about the problems and difficulties that lie behind exam-taking in general and language exams in particular, in the Greek context, chooses to give his article the title ‘Exams - The Necessary Evil’ and stresses the unfairness of assessment based on a single performance and how the amount of luck involved makes the exams take on the nature of a lottery. Sifakis and Sougari (2010, p. 305) point out that FLL in Greece is considered “crucial” due to the fact that Modern Greek is not widely-used outside Greece, with English being the “primary foreign language” of preference. They note that in the last three decades there has
been an increasing demand for “certification of proficiency in English”, for which young people “prepare strenuously”.

Broadfoot (2005, p. 125) makes the very valid point that schoolchildren, through being subjected to summative assessment processes and high-stakes testing are being put through an inordinate amount of pressure that most adults would be unable to bear. Prodromou (1995, p. 14) emphasises the fact that Greek society is one in which “examinations play a significant role”, with the result that the kind of atmosphere in the EFL classroom tends to be one of a “stressful textbook-bound environment” and he warns that the “backwash” effect resulting from the pressure of public exams can be both “overt” and “covert” and that materials which are unlikely to feature in the exams, will, however useful to the learner, rarely be “brought up” in class. In an interview with Spyropoulou (2006), George Vassilakis, a veteran teacher and teacher-trainer, also highlights how language teaching and learning in Greece, and the coursebooks used in this process, tend to minimise the importance of learning and over-emphasise the importance of exam-taking, which he finds disturbing.

Fabian (1982, p. 24) goes so far as to describe examinations as a “formal and unnatural ritual” and there is no doubt that public EFL examinations in Greece have had a far-reaching influence on Greek society and have played on the not ‘unnatural’ desires of parents for their children to have validated certificates attesting to their level of knowledge of English and other foreign languages. Fabian also claims that these examinations “thrive on faith” and in the case of Greece, this degree of “faith” has resulted in a huge, flourishing market for examination boards such as the University of Cambridge, the University of Michigan, Edexcel and many more. Fabian (ibid) is also of the opinion that examinations “influence to an unacceptable and unreasonable degree” and many academics (Freeman, 1995; Stefani, 1998;
Clifford, 1999; Biggs, 1999) believe that at the tertiary level, at least, the way in which students are tested, influences to a large extent what and how they learn.

Interestingly enough, Shephard (1982, p. 28) also believes that language exam candidates are “full of faith” and he makes the very condemnatory statement that “examinations stultify and constrict, emasculate the new and perpetuate the old, and appeal to the lowest motives on the part of all concerned”. In making this statement, it could be said that Shephard touches on four areas which are of relevance to the present research and these are:

1. the effects which TEA have on language learning and on learner autonomy;
2. the resistance to change on the part of teachers, learners and society at large;
4. what Biggs (1999, p. 61) refers to as students “watching their backs”.

The first point will be discussed further in this chapter, while the other matters are clearly deserving of further investigation and will be referred to in the Discussion chapter.

### 3.1.3 Who controls TEA?

In the Greek educational system as in many educational systems throughout the world, the matter of TEA has been dominated and controlled by ‘authoritative others’ and in Greece, this type of TEA continues to reign relatively unquestioned in all spheres of education: primary, secondary and tertiary. Assessment in language education at these levels is no different from assessment of other subjects in the curriculum, and is usually based only on written ‘end-of’-week, -month, -term, -year, -course tests in order to meet the criteria and standards set down by the Greek Ministry of Education.
The same attitudes to TEA prevail even in private language schools (φροντιστήρια) where the main interest of students and their parents is to acquire certification and attestation as to their knowledge of English and other foreign languages by means of sitting the public exams offered by the previously-mentioned recognised examining bodies. Preparation and coaching for these exams, with a few exceptions, rarely requires students to make judgements about the language performance of themselves or their peers and they depend on the teacher, and ultimately, the examining body, to do this for them.

3.1.4 TEA and its influence on learner beliefs and attitudes
While the majority of the participants in the AARP have devoted a great deal of time and effort, and, of course, money, to gaining certification in English, usually in the years prior to the two years of intensive preparation for university entrance, and many have reached a high level of achievement, such is their blind faith in their abilities (and they have the certificates to prove it) that they are quite unaware that time and lack of practice have slowly eroded their previous knowledge, and, in effect ‘shrunk’ their certificates, making it not uncommon for students to have fallen as much as two levels on the CEFR scales (Everhard, unpublished data). Just as they have forgotten a good deal of what they learned in order to enter university, in the same way the technique of memorising in order to pass language exams has taken its toll. Nevertheless, students can be inordinately stubborn about accepting these facts, even when presented with supporting evidence. In the majority of cases, their engagement with learning has been ‘surface’ rather than ‘deep’, enabling them to demonstrate enough knowledge to successfully pass examinations, but not to become true owners of that knowledge (Biggs, 1999; Clifford, 1999; McKay & Kember, 1997; Somervell, 1993; Stefanou et al., 2004; Yorks & Kasl, 2002).
3.1.5 TEA and its influence on learner autonomy

Little (2003, p. 225) deals with the topic of public examinations and the “powerful washback effect” that they have and how they can be counteractive in fostering autonomy; however, he feels that there is hope for the future through the work achieved by the Council of Europe (Little, 2003; Little, 2005a, p. 324) and states that:

perhaps between them the CEFR and the ELP can contribute to the development of an assessment culture in which self-assessment can help to bring the learning process into a closer and more productive relation to tests and examinations than has traditionally been the case.

The Council of Europe (2001, p. 192) believes that self-assessment combined with other methods of testing and evaluation is useful in situations where ‘high-stakes’ testing is not involved and states:

the main potential for self-assessment, however, is in its use as a tool for motivation and awareness raising; helping learners appreciate their strengths, recognise their weaknesses and orient their learning more effectively.

A very similar view is taken by Oscarson (1989), while Ekbatani & Pierson (2000, p. 152) stipulate that self-directed assessment is “a valid and reliable addition and supplement” to more traditional approaches to TEA, which they believe can provide teachers with a “more accurate picture” of what learners are able to do.

Like Little (2003), Raappana (1997, pp. 131-132) feels that traditional methods of assessment have a lot to answer for and can work against autonomy, since “often it is the grading system that makes attempts to enhance self-direction futile at school”. She also highlights the strong temptation that drives an exam-oriented learning community, when she states that “the tradition of ‘teaching for the examination’ is hard to break”.

The other problem that is encountered, she feels, is in convincing others that the previous monopoly that teachers had over assessment should be abandoned, since
“there is still a very strong tradition to regard the teacher’s assessment as infallible, unquestionable, which is natural considering the tradition of assessment based on the ‘closed input-output circle’” (Raappana 1997, p. 132).

Little (1990, p. 12) observes that the obsession with exam-taking and grades can be detrimental to the uptake of autonomy and to changing traditional learner-teacher roles, since:

[a]utonomy – accepting responsibility for their own learning – may be the last thing learners want. In many cases their chief interest is in acquitting themselves well in exams, and it may be difficult to shake their belief that the teacher’s job is straightforwardly to prepare them for that end.

Little (1999a, p. 33) makes clear why such beliefs prevail, stating that:

[u]nfortunately most public and state examinations are still shaped by a tradition which assumes that the purpose of education is the accumulation of facts rather than the development of skills, and which places more emphasis on outcome than on process.

Little (2003) warns us that washback from public examinations can have a stultifying effect on autonomy, but in outlining what distinguishes formative assessment from summative assessment, Little (1999b, p. 3) also warns us of another possible washback effect, which is that in using formative assessment, it is possible that it could result in following the “same procedures” as with summative assessment. While Little is not altogether explicit about what these “same procedures” might be, it is clear from what he says previously that the warning he intends to give is that:

(1) formative assessment might come to be used at the end of a learning phase in the way that summative assessment has traditionally been used, so that the grade awarded would remain what is all-important to learners;

(2) learners might not be provided with timely and useful feedback in order to learn from the process and experience of self-assessment, rendering it, like summative
assessment, assessment of learning rather than assessment for learning (Earl, 2003; D. Little, personal communication, November 26, 2009; Stiggins & Chappuis, 2006; 2008). Paris (1998) and Dann (2002) also remind us that the use of assessment for accountability purposes has been exploited to the detriment of improved learning and increased motivation.

With regard to evaluation and assessment, Dam & Little (1998, p. 129) highlight the fact that practice is necessary in order for learners “to become good decision makers”, in the same way that they need practice in order “to become good communicators in their target language”. Kohonen (2001b, p. 53) is also of the opinion that by increasing the learners’ “capacity for self-assessment” we facilitate autonomous language learning, but he likewise believes that learners require “time, guidance and encouragement” in order to acquire that capacity.

This can be difficult to achieve, given the long-standing beliefs and attitudes of learners and teachers towards TEA. Chen (2006a, p. 14) asserts that students may not wish to be active participants in the assessment process, while, likewise, Black, Harrison, Lee, Marshall and Wiliam (2003, p. 98) have evidence that learners who have played a “passive” role in the learning process may actively resist attempts to change that. Little (1990, p. 12) points out that attempting to change learner beliefs can be very disconcerting and “we should not be surprised if some learners are resistant to autonomy; for autonomy implies a continuous challenge to our certainties, and that can be very unsettling”.

Allwright (1988, p. 35) asserts that for many years, terms like ‘autonomy’ were so far removed from mainstream pedagogical trends that they were considered to be ideas belonging to “the lunatic fringe” or were, at best, some kind of “exotic plant” that it was generally considered could not possibly be “cultivated” in the traditional
language classroom. It is therefore not surprising that attempts to cultivate autonomy through changing assessment practices might meet with resistance.

3.1.6 The need for assessment literacy
Previously, the warning from Little (1999b, p. 3; personal communication, November 26, 2009), that there is a danger of formative assessment being put to the wrong use and being used in a purely summative manner, was mentioned. This means that the instructor, due to lack of preparation or training, is not sufficiently literate as regards assessment to be able to put it to good use. Likewise, there are other ways in which formative assessment can be abused or misused, which should not be forgotten. At this point, it seems appropriate to quote the comments of a seminar participant who is cited by Dickinson (1987, p. 134) as follows:

The very idea of self-assessment is a contradiction in terms. How can an individual learner be expected to take an objective view of his or her own attainment without succumbing to the temptation to cheat? The whole idea is ridiculous!

Indeed, Gardner (2000, p. 54) makes the point that any kind of assessment is only of value if it is accepted as valid, so that awareness of the usefulness of self-assessment has to be raised in all quarters if its “face validity is not to be compromised”. Issues such as the learners’ temptation to cheat, resistance to new responsibilities, or disrespect towards teachers due to what seems like the offloading of work, as well as teachers’ loss of power, or worries concerning the potential threat to their jobs, all have to be dealt with and resolved. Gardner feels that it is the job of teachers to ensure that “learners acquire the requisite skills”.

Rolheiser and Ross (2003), writing in an American context, likewise also point out that winning students over to the idea of self-assessment and dealing with their misconceptions and pre-determined views is a challenge. Not only that, but teachers
also have to accept “the recalibration of power that occurs” (Rolheiser & Ross, 2003, p. 13) when judgement concerning assessment is shared with learners. This means not only that the learners have to become assessment literate, but that teachers also have to experiment in order to enhance their own assessment literacy. All of this involves changing “root beliefs, behaviours and relationships” which Rolheiser and Ross (2003, p. 13) admit is both difficult and takes a lot of time. However, with what seems to be a move in schools and institutions of higher education towards creating a more democratic environment, with sharing of leadership and responsibility for learning, it becomes clear that “traditional assessment practices are no longer adequate” (Rolheiser & Ross, 2003, p. 3) and that the school or academic environment and “curriculum-assessment experiences should prepare students for life in the real world” (Rolheiser & Ross, 2003, p. 1).

Likewise, Hirvela and Pierson (2000, p. 105) feel that it is the responsibility of teachers to understand and realise when to move beyond summative to more formative forms of assessment and that “teachers may best serve students and themselves by knowing when to step back from traditional approaches to assessment and allow for learner-directed assessment”.

Such a decision, they feel, has to be grounded in the conviction that, firstly, students are able to measure their learning achievements in a meaningful way and, secondly, that a reliable form of self-assessment can be created in a complex area or skill, such as that of writing. What Hirvela & Pierson (2000, p. 108) feel is the most important outcome of formative approaches to assessment, involving self-assessment, is the fact that it involves the learner in being “an active and creative participant in the assessment process, rather than an object of it”. This point of view would seem to coincide with that of Kohonen (2006a, p. 28) and Little (1999b; 2005a) who, in
connection with the ELP, assert the importance of self-evaluation and self-assessment, with Kohonen (ibid) stating that “pupils must also be able to evaluate their learning outcomes and identify their weaknesses as well as their strengths in order to set new aims for the next cycle in their language learning progress”.

3.2 Alternative assessment, the COE and the CEFR
Interest in alternative forms of assessment which would identify more with the Council of Europe’s policies on language education and the promotion of autonomy, was initially aroused in language education circles through the work of Oskarsson (later spelled Oscarson) and, in particular, through the two Council of Europe documents he produced on self-assessment in 1978 and 1984, respectively. Both documents were surveys of what was happening, in terms of research and practice, in relation to self-evaluation in language learning, in a number of European countries and beyond.

Within the Council of Europe, the publication of Oscarson’s work coincided, to some extent, with that of van Ek (1977) and Trim (1978) on what were initially called the European Threshold Levels and with the publication of the work of Holec ([1979] 1981) on the promotion of autonomy in language learning.

In more recent years, within Europe, it has been the introduction of the Common European Framework of Reference (CEFR) (Council of Europe, 2001) and the European Language Portfolio (ELP), in 1997, and their development, which have both played a key role in forming and reshaping attitudes towards testing, evaluation and assessment (TEA) and in ensuring that learners play a more pro-active role in their learning, rather than simply a reactive one. Two of the key players in developing the CEFR were John Trim and Brian North, while, in the case of the ELP, David Little was and continues to be one of the prime ‘movers and shakers’. It could be the case
that his work is driven by the conviction (Little, 1990, p. 12) that “it will be the autonomous learners who most easily make the transition from learner to learner/user of the target language”.

Little (1999b) refers to the tremendous influence that the work of the Council of Europe (COE) has had with regard to the compilation of language syllabuses, curricula and the creation of learning materials, but felt that at that point in time that the COE’s impact on assessment had been less far-reaching and also “less complete”. Part of the problem with regard to this reluctance or slowness to change, he felt, was the fact that forms of assessment in schools and universities had also been very slow to change, as these still tended to be traditionally competitive, and therefore contrary to the criterion-referenced approaches promoted by the Council of Europe, which focus much less on failure and more on how far learners have mastered a particular skill or behaviour. The CEFR has been fundamental not only in formulating common standards for all the European languages and their assessment, through bodies such as the Association of Language Testers in Europe (ALTE), but has also become the fundamental reference point for teachers and for learners on progress made in the foreign-language learning setting.

Indeed, Little (2006) makes it clear that the function of the CEFR is not just to guide testing and examinations of progress with foreign languages, but is also to form a guide and source of inspiration for language syllabuses, curricular guides and language coursebooks. While the importance of reflection and self-assessment in the ELP pilot projects is underlined in Little (2002), the importance of learner autonomy is highlighted in Little (2006) as being essential to the use of self-assessment checklists and the learner-centred curriculum of the ELP.
Many publications have been released at a national and pan-European level under the auspices of the Council of Europe and the European Centre for Modern Languages and various documents have been produced which guide teachers and developers, at a national or a regional level, in the use of the CEFR and the ELP, of which the three constituent parts are the Language Passport, the Language Biography and the Dossier. It becomes clear from even a summary perusal of any one of these documents that the CEFR forms the backbone of the whole project.

What is of significance to the AARP is the fact that there has been a very clear shifting in the teaching of foreign languages within Europe to the promotion of autonomy, through reflection, which is engendered through self-assessment of progress, by means of grids, check-lists and portfolio documentation. Indeed, during the academic year 2010-2011, learner-centred assessment has been introduced in foreign language learning at primary and secondary levels, in Greece, and there are plans to introduce project work and self-assessment of that project work in a new subject to be piloted, in the academic year, 2011-2012, in the curriculum in the last year of senior high school. This is an attempt to break away from the traditional methods of rote learning and to encourage critical analysis and greater autonomy. Like the AARP, such a move could be the beginning of something much bigger.

In the sections which follow, we will take a look at the researchers Holec, Oscarson, Dickinson, Little and Kohonen, who have been the main contributors to thinking about autonomy and assessment in EFL within Europe. The various ways in which each has added to our understanding of the relationship between the two will be considered.
3.2.1 The contribution of Holec to the field of assessment

Although Holec views evaluation as being one of the prime objectives of the language learner, both through monitoring and assessment, his writings on the subject, based on work which was conducted in his workplace at CRAPEL in the University of Nancy, France, in the 1980s, unlike his legendary definition (Holec, 1981) are rarely, or almost never referred to. This is indeed unfortunate as careful and close reading of the few texts which Holec has published on the subject provide many insights and it seems likely that much of the confusion which surrounds autonomy and its promotion in language learning could have been avoided if researchers in the field had paid greater heed to: (a) the context in which Holec saw the promulgation of autonomy being possible; (b) the key role that Holec saw self-evaluation/self-assessment as playing in that promulgation.

For Holec (1981, p. 9), the five constituent elements enabling the learner “to take charge of one’s own learning”, are:

1. determining the objectives;
2. defining the contents and progressions;
3. selecting methods and techniques to be used;
4. monitoring the procedure of acquisition;
5. evaluating what has been acquired.

In fact, Holec (1981; 1985a; 1985b) sees a very close relationship between objectives and evaluation and therefore believes that the criteria by which the learner’s progress will be judged should be chosen by the learners themselves, with the criteria and their combination varying in relation to the learners’ objectives and their current level of knowledge. Indeed, Holec believes that self-evaluation plays an even more significant role in developing the learner’s self-directedness than the setting of objectives. Henner-Stanchina & Holec (1977, p. 75)) go so far as to claim that there can be no learning without evaluation as it is both an “integral” and “internal” part of the
learning process, helping the learner decide how to continue or modify their learning activities; thus, it is necessary both during each step as well as at the end of each step in the learning process.

This process of evaluation is “entirely different from traditional external evaluation” since it is not of a comparative nature, it is not norm-referenced and it is not other-referenced (Henner-Stanchina & Holec, 1977, p. 79). This therefore means that learners at CRAPEL simultaneously create and undergo the evaluation procedure, judge their achievement in relation to themselves, judge themselves against their own personal criteria and judge themselves in accordance with their own objectives and learning expectations.

Holec (1985a, p. 11) anticipates that learners may encounter some problems which could present themselves on a social level, since this kind of behaviour might not be “compatible with the way he behaves in other circumstances”, and, on a psychological level, because it requires “destruction/reconstruction”, “self-criticism” and coping with “fear of the unknown”. Teachers may also encounter problems due to existing prejudices (Holec, 1985b, p. 141) concerning self-assessment, such as:

1. Learners are unable to assess themselves;
2. They simply can’t be trusted to do it;
3. Only someone who ‘can do something perfectly has the right to judge that performance’;
4. Learners themselves have no confidence in self-evaluation;
5. Evaluation is an instrument of power which should not be put in just anyone’s hands.

Such beliefs are frequently held by the learners themselves and it will require a great deal of effort to change and “modify the behaviours of the participants involved” (Holec, 1985a, p. 7). Holec (1985a, p. 18) sees autonomy as being an area “full of
ambushes and pitfalls”, but, clearly, thinks the risks are worth taking when he asserts, “not to be careful would be foolhardy, but not to join in would be cowardly”.

### 3.2.2 The foresight of Oscarson (formerly Oskarsson)

Oskarsson’s first COE publication (1978, p. 3) shows him as a man of great foresight with regard to the yet-to-be European Language Portfolio, when he stated that self-assessment would become “an increasingly important feature of the system”, creating a need for “evaluation techniques which can be put in the hands of the learners”.

Oskarsson believes that self-assessment holds promise in two respects. Like Holec, he sees self-assessment techniques as useful in providing “individual continuous feedback on what has been learnt” and in assessing “total achievement at the end of a course or course unit” (author’s emphasis).

Moreover, Oscarson (1989, pp. 2-3) emphasizes that research results from within Europe indicate that “the validity of learner judgements can in fact be quite high” and he regards the implementation of self-assessment as justified because it brings “widened perspectives” to both learners and teachers so that they come to view assessment as “a mutual responsibility, not as the sole responsibility of the teacher” (author’s emphasis), which he views as being “conducive to the democratic development of language teaching”. Like Kohonen (1999), he believes it can engender positive affect.

Oscarson (1989, p. 11) clearly sees a strong link between self-assessment and learner autonomy and feels that more emphasis should be given to “the importance of giving students practice in autonomous learning and self-directed evaluation, at all levels and in a wide variety of language learning settings” (author’s emphasis), since learners are “unaccustomed to the very notion of self-reliance”. In order for this to
happen, Oscarson reminds us of the fact that teachers have to receive the training that is necessary to put this into effect.

Among the many advantages that Oscarson (1997, pp. 175-176) sees in self-assessment are, that:

1. Learning is more effective and relevant, engaging the learner in all phases of the learning process;
2. The reflection process helps in building new knowledge;
3. Assessment of learning based on internal experience will benefit learning;
4. It will have positive effects on learner autonomy, motivation, learning outcomes and learning skills;
5. Cooperation and partnership will be created between learners and teachers;
6. It will promote lifelong learning by developing a lifelong skill.

Oscarson (1997, p. 184) believes that learners go through three distinct stages in their need for support and in their progress as assessors:

- **a. Dependent stage: Full dependence on external assessment**
- **b. Co-operative stage: Collaborative self- and external assessment**
- **c. Independent stage: Full reliance on independent self-assessment**

![Figure 3.1 Reflection and assessment on the heteronomy-autonomy continuum](image)

These three stages, as we can see in Figure 3.1 above, correspond, in their progression, with Huttunen’s (2003, pp. 125-126) three stages of reflection, fitting perfectly into the continuum and model of learner autonomy outlined in Chapter 2.
Like Von Wright (1993, cited in Oscarson, 1998, p. 2), Oscarson believes that a significant role in the learner’s development is played by the “internal feedback” generated by self-assessment since it helps the learner to engage in learning at a deeper level, leading Von Wright to declare that “(w)ithout the ability to evaluate one’s own action(s) learning takes place at a superficial level”. Oscarson (1998) also agrees with Lewkowicz & Moon (1985, p. 46) that assessment is “a powerful weapon of control” and he emphasises the “natural need of learners to be in control of some aspects of the evaluation of their own learning” (Oscarson, 1998, pp. 8-9), which he thinks should be combined with “dependable and unbiased language testing procedures in high-stakes decision contexts”. In this way, through the common endeavours of teachers and learners in the assessment process “better conditions for language learning” will be created, but Oscarson and Apelgren (2011, p. 3) feel that much more research conducted using alternative approaches, such as portfolios and the use of peer-assessment and self-assessment, needs to be reported.

3.2.3 Dickinson’s contribution to the field
Few instructors or researchers in the field of Applied Linguistics would argue with Dickinson’s (1987, p. 26) comment that “(t)here are few tests that assess increase in learning rather than simply count correct answers” and he views self-assessment as being a very important way of countermanding that.

Dickinson has been marginalised to some extent because he prefers to refer to ‘self-directed learning’ or ‘self-instruction’, which others take to mean learning in isolation, rather than autonomous learning. This is unfortunate because Dickinson has some very enlightening comments to make about TEA in relation to autonomy and both his monographs (Dickinson, 1987; 1992) place emphasis on self-assessment, devoting quite a large amount of space to the subject. Dickinson (1995, p. 166) sees
autonomy as “an attitude towards learning” as well as “a capacity for independent learning”. In order for this “attitude” and “capacity” to be fostered and developed, Dickinson feels very strongly that teachers have to convince learners that:

1. judging the degree of success of their learning is “an essential ingredient”;  
2. personal involvement in decision making “leads to more effective learning”.

Dickinson (1987, p. 16) believes these are essential in encouraging autonomous learning.

In Chapter 2, Table 2.5 the areas of consensus on autonomy were outlined, three of these being that autonomy is “multi-faceted”, “varying in degree” and is a “process” rather than a product. Dickinson (1987, p. 137) thinks that the same is true of assessment. He also believes that there may be many different degrees of “the self” in self-assessment, so the concept warrants thorough analysis. Gardner (2000, p. 54) sees self-assessment as taking its validity from the attitudes of the parties involved and Dickinson (1992, p. 36) likewise believes that self-assessment becomes “legitimized” only by persuading the learner that it is “a valid and useful activity”. Language teachers have the added responsibility of not only helping the learner progress with the language, but enabling her to “become more proficient in self-assessment”, so she “can become progressively better at judging her own performance” (Dickinson, 1992, p. 32). He also believes that discussion and analysis which helps the learner understand in what way their performance does not reach the required standard is likely to lead to greater learning. Moreover, he sees the heightening of learner self-assessment as naturally reducing the need for teacher involvement, and, in this way, building “greater empathy between teachers and learners” (Dickinson, 1987, p. 26). Dickinson views record-keeping as being an
essential element in self-assessment in order to enable the learner to recognise progress over time.

Dickinson (1992, p. 35) asserts that there are basically two obstacles that may interfere with learner self-assessment. One is the learners’ lack of belief and confidence that they are capable of self-assessment and the other is their unwillingness to recognise and accept self-assessment as “a legitimate activity”. It is therefore the duty of the teacher to provide the learners with opportunities for “supported practice”, which Dickinson suggests can be achieved through the exploitation of what he calls “frozen data”, which can either be taken from the students’ past learning experience or from students at a lower level (why they should be at a lower level is not explained).

Dickinson (1992, p. 34) provides us with a very convenient and analytical list of the qualities required of a self-assessor and the demands and responsibilities placed upon them. Self-assessment may involve:

- the willingness and motivation to undertake it
- the willingness to reject inadequate performance (rather than merely to shrug one’s shoulders in indifference)
- some internal standard, either established by oneself or learned
- the ability to measure one’s performance against this standard
- the confidence to make these assessments
- the recognition that one’s ability to judge, and the accuracy of the judgement may be very limited.

Finally, Dickinson (1987, p. 150) is brutally honest about the possible drawbacks of self-assessment and informs us that it is not a matter of whether learners “can” self-assess, but rather a question of whether they “will” self-assess. In addition to the ‘willingness’ to take responsibility for their learning, he also means that what is
brought into question is whether the learners will be totally honest in their judgements or whether they will “succumb to the temptation to cheat”.

To this he replies that when it is a matter of scores and grades being put on display to others, then the temptation to cheat will be high. When it is a more personal case of the learner recognising whether learning objectives have been achieved, then Dickinson feels that “cheating becomes pointless”. Perhaps this is why Dickinson (1992, p. 35) believes that peer-assessment is a useful training ground for learners in preparation for self-assessment.

3.2.4 Little’s Council of Europe work

Little’s deep involvement in the development and deployment of the ELP at both national and European levels means that he has thoroughly investigated self-assessment from all aspects: theoretical, practical and philosophical. Little (2005a, pp. 321-322) offers us three reasons why we should engage our learners in the practice of self-assessment and we should pay attention to the outcomes. These are outlined as follows:

(1) …a learner-centred curriculum…falls short of its definition if learners are involved in decisions regarding the content of the curriculum and how it is taught but excluded from the process of evaluating curriculum outcomes, including their own learning achievement.

(2) …making self-assessment an integral part of the evaluation procedures not only encourages learners and teachers to regard assessment as a shared responsibility, but it also opens up wider perspectives on the learning process.

(3) …to the extent that languages learnt in formal contexts are to be used in the world beyond the classroom, a capacity for accurate self-assessment is an essential part of the toolkit that allows learners to turn occasions of target language use into opportunities for further explicit language learning.
Little (2009a, p. 1) sees “the primary orientation” of the CEFR as being “behavioural” and views the descriptors provided in the CEFR as being capable of working on three different planes, so that they can be used to:

1. define a curriculum;
2. plan a programme of teaching and learning;
3. guide the assessment of learning outcomes.

Little feels that this is remarkable because “the CEFR offers to bring curriculum, pedagogy and assessment into a closer relation to one another than has traditionally been the case, challenging us to rethink each from the perspective of the other two”. Unfortunately, despite this extraordinary ability, Little (2009a, p. 2) believes that the CEFR and ELP have not had the “impact on language learning outcomes” that might have been expected, because they have tended not to be “exploited as a single package”.

Little & Perclová (2001) describe self-assessment as being “central” to the ELP, with its three parts of Language Passport, a Language Biography and a Dossier. Self-assessment underlies both the Passport and the Dossier and encourages a sense of ownership, which, in turn, promotes lifelong learning. In pilot ELP projects undertaken in various European countries, positive reports on self-assessment came back from teachers, who claimed that it helped open up dialogue with their learners, which made the teachers more appreciative of their learners’ problems, while learners reported satisfaction at having control over their learning.

Of course, as with anything that deviates from what is known and well-tried, some reservations were expressed, such as the following three points, outlined by Little (2009a, p. 3):
Learners do not know how to assess themselves. There is a danger that they will overestimate their proficiency. They may be tempted to cheat by including in their ELPs material that is not their own.

As was only to be expected, some doubts and disbeliefs were expressed concerning learners’ ability to self-assess, but since self-assessment it so central to the functioning of the ELP, gradually, through practice, such doubts are put aside. Indeed, the scales, levels and descriptors set out in the CEFR are put into effect.

With regard to honesty, like Dickinson (1987; 1992), Little (2009a, p. 3) suggests that “if ELP-based assessment is central to the language learning process, there is no reason why it should not be accurate, reliable and honest” and it is self-assessment which is “the dynamic that drives reflective language learning”.

Clearly, in the case of the ELP, creating a culture of self-assessment is not easily achieved but is mandatory for its success. The ELP embodies an altogether different philosophy towards achievement and failure compared with traditional TEA since its criterion-referenced assessment promotes positive attitudes. Formative assessment is used all the way through the course to provide learners with feedback. Although self-assessment is associated with formative assessment, sometimes summative assessment is required of learners at the end of a phase of learning or a course.

What is of particular relevance to the AARP, on which the present thesis is based, is the fact that “because self-assessment in the ELP is based on the common reference levels of the CEFR, self-assessment, teacher assessment and external assessment can all orient themselves to the same behavioural descriptions” (Little, 2005a, p. 323). Little (1996b, p. 31) sees self-assessment as an “indispensable element” in the development of strategic control of the learning process and of
language use, but emphasises that this is achieved through “instructional conversation” as well as through “negotiation” and “mutual support”.

Little (2005a, p. 1) feels confident that the CEFR, in combination with the ELP, will effect the changes that are necessary to overcome traditional ideas and beliefs concerning approaches to learning and TEA. He feels that together they will provide the opportunity to “renew the challenge to develop a culture of assessment that both facilitates and takes full account of learner self-assessment”.

Little & Perclová (2001, p. 53) recommend peer assessment as a useful springboard for self-assessment, citing the two reasons that, firstly, it is easier to detect faults and errors in the work of others and, secondly, one learner’s knowledge is not the same as another’s and so one learner may be able to offer advice to another based on what they know. Little & Perclová also suggest that if the same criteria are used by peers as will be used in exam conditions, then this process may actually assist them in their own exam performance.

In evaluation of task performance, Little states elsewhere (1996b, p. 31) that self-assessment interacts “inevitably” with “other” or “peer” assessment. Little goes so far as to say that “classrooms in which self-assessment interacts fruitfully with peer-assessment have probably gone as far as it is possible to go in the promotion of learner autonomy”. This assumption forms the crux of investigations in the AARP.

3.2.5 Kohonen’s experiential approach
Another voice within Europe, which perhaps has not been heard as clearly and loudly as it could have been, is that of Kohonen. Kohonen is worthy of our attention because he places autonomy and evaluation/assessment within a framework of what he refers to as experiential learning (1990; 1992a; 2001b) and, equally importantly, he also
places language learning within the broader notion of what he calls (2007, p. 4) “learner education”, which he summarises as involving the following:

1. The student’s own goals and autonomy;
2. Personal engagement in learning;
3. Student initiative and responsibility;
4. Meaningful learning as a whole person approach;
5. Emphasis on reflection, interaction and self-assessment;
6. Integration of social and affective learning with cognitive goals.

Kohonen (2002, pp. 1-2) views language learners and language teachers very much as members of what he calls a “collaborative learning community” where through shared practice and cooperation, each can develop fully through “a deliberate shift towards collaborative, active and socially responsible learning in school”. For Kohonen (1992b, p. 74), it is clear that assessment plays a key role in developing learner autonomy when he asserts that “a fully autonomous learner is totally responsible for making the decisions, implementing them and assessing the outcomes without any teacher involvement”.

The feelings and attitudes of participants are also something that he sees as being of importance. Indeed, Kohonen (1992a, p. 23) goes so far as to claim that “the affective component contributes at least as much as and often more to language learning than the cognitive skills”. Kohonen (ibid) believes that confidence is gained through competence and that competence, in turn, is acquired through “internalization of the criteria for success”, which he thinks is “fostered by teaching that encourages the learner’s self-assessment of his or her own learning, both alone and with peers in cooperative learning groups”.

With regard to evaluation and assessment, Kohonen (1999, p. 282) warns us of the danger of contradictory practices in teaching and testing, from both the teacher’s
and the learner’s perspective. While teachers may well be working towards “learner-centred classroom practices” on an everyday basis, when it comes to testing, it could be that the procedures that they follow may be very tightly “teacher-controlled”, which, as he points out, would undermine their “authenticity as a learner-centred educator”.

From the learner’s perspective, while Kohonen (2000; 2006b), is a great proponent of the ELP and the use of portfolios and portfolio assessment, he reminds us also (1999, p. 289) that not all learners are alike and therefore may not all be willing to make this shift, citing the example, in one pilot project, of 14 boys who, unlike their peers, found portfolios “distressing”, “demanding” and “requiring too much work”. For this reason Kohonen suggests that “we need to be sensitive to diverse learner needs, beliefs and expectations and see how far self-assessment is viable for the different learners”. He likewise suggests that learners need a great deal of “encouragement”, “explicit guidance”, “support” and, above all, “time” if they are to adopt and learn the “new skills of self-assessment”.

In general, though, Kohonen (2006b, p. 18) views working with the ELP as encouraging and claims that the Finnish ELP project furnishes evidence that the “involvement” and “engagement” which working with the ELP fostered, encouraged feelings of “ownership” amongst Finnish learners, since “it helps to make the learning processes more concrete and observable”.

In line with the ELP, Kohonen (1992b, p. 83), as a great believer in Portfolio Assessment, asserts that evaluation should be both “process-oriented”, since process evaluation helps both the teacher and the learner gauge the progress that is being made, and “product-oriented”, which as a form of summative assessment, is
concerned with “levels of attainment”, while “process evaluation” can be used more in “shaping” the “learning process”.

The reflective processes involved in authentic assessment help to build learners’ awareness and to form their identity as persons as well as language learners and through “contextualised and specific feedback” helps to inform them of the progress that is taking place (Kohonen, 2001a, p. 35).

This concludes the overview of the most influential figures in Europe regarding autonomy and assessment in EFL. The next section will look at two non-EFL researchers on assessment, with a more global perspective.

3.3 Two world-leading proponents of self- and peer-assessment

Before moving on to the research studies most relevant to the AARP, it is necessary to look, albeit briefly, at the work of two figures whose work has not been confined to the world of EFL or to work within the council of Europe. Indeed, their writings on self-assessment and peer-assessment have been not only prolific, but have been extremely influential worldwide. The two figures referred to are Boud, most of whose work has been concerned with self-assessment, and Falchikov, most of whose work has been concerned with peer-assessment. In recent years, however, they have cooperated with each other, giving the work of both a much wider perspective. We will first look at the work of Boud.

3.3.1 Boud, autonomy and self-assessment

One of the first academics, it would seem, to make the essential link between the fostering of autonomy and assessment was Boud (1981, p. 25), who sends out a powerful message concerning the importance of encouraging autonomy in students in higher education, although he admits it may not be easy and fears that
“[p]ostponement of the opportunity to exercise responsibility for learning actively discourages the development of the capacity to do so”. 12

Boud (1981, p. 25) sees decision-making as being at the heart of autonomy and believes that this decision-making process should include the participation of learners in “crucial” areas such as assessment. He does not believe that teaching which constantly places the teacher at the centre of activity can be conducive to autonomy.

The bulk of Boud’s academic output, however, spanning decades, relates to assessment and its promotion in HE. What Boud (1995, p. 11) regards as being the two key elements which assessment in HE promotes are:

(1) … the development of knowledge and an appreciation of the appropriate standards and criteria for meeting those standards which can be applied to any given work.

(2) … the capacity to make judgements about whether or not the work involved does or does not meet these standards.

Boud (ibid) feels that there are genuine opportunities for learning, in the first case, which are usually ignored in favour of the second. Self-assessment should involve much more than just self-grading, he feels, but should involve the learner in the deeper processes of being able to decide objectively what constitutes a good piece of work in any particular case. This ability, which takes students “beyond the present context”, contributes to a learner’s development through increased “self-knowledge” and “self-understanding”, which Boud (1995, p. 20) sees as being “emancipatory”. He views self-assessment as therefore providing “the fundamental link with learning” (1995, p. 15).

Thus, Boud asserts that when we use the term ‘self-assessment’, we have to be very clear as to exactly what we mean by it. Boud (1995, p. 13) refers to self-

12 The italics from the original are preserved.
assessment variously as a “process”, an “activity” and “a goal to which to aspire as well as a practice in which to engage”. While Boud points out the very strong links between reflection and self-assessment, since “[a]ll self-assessment involves reflection”, at the same time he emphasises that “not all reflection leads to a self-assessment” (Boud & Brew, 1995, p. 131).

Boud earnestly believes that students miss out on a golden opportunity for learning if they are not involved in the discussions about or in the setting of criteria for a particular piece of work, although he acknowledges that in some cases a lack of expertise or sophistication on a given topic could preclude them from deciding on criteria to be set.

Self-assessment and university
Boud (1995) is convinced that self-assessment is a necessity for “effective learning” and so it is something that must be encouraged as part of a university education and this will stand students in good stead for lifelong learning. Thus, Boud (ibid) considers that the outcome of a university education should be that students can continue to learn without resort to course-driven or teacher-driven assistance or without dependence on supervisors for assessment in the work-place. Boud (1995, p. 14) outlines the four advantages to be gained by graduates due to their involvement in self-assessment, which are:

- They are more likely to wish to continue their learning.
- They are more likely to know how to do so.
- They are more likely to monitor their own performance without constant reference to fellow professionals.
- They are more likely to expect to take full responsibility for their actions and judgements.
Boud believes that engagement in self-assessment enables learners to draw “meaning from experience”. However, at the same time, we must accept that an individual’s judgement will be coloured by all their past experience up to that point, so that the same event will be viewed in different ways by different individuals.

What is very interesting is that Boud & Brew (1995) have taken the same “three knowledge constitutive interests” of Habermas, as has Huttunen (see Chapter 2, Section 2.3.2) and linked them with three levels of self-assessment, which is almost an exact match with Huttunen’s (2003) three “interests” for reflection and autonomy; however, while Huttunen (ibid) refers to 1) Mechanical reflection; 2) Pragmatic reflection and 3) Emancipatory reflection leading to autonomy, Boud & Brew (1995) refer to 1) Technical interest; 2) Communicative interest and 3) Emancipatory interest, leading to Self-assessment. Thus, we can build on the previous model, adding Boud & Brew’s (1995, pp. 133-134) and Brew’s (1999, p. 165) suggestion and create a more complete view of the autonomy continuum in Figure 3.2 below:

![Figure 3.2 Stages of reflection and assessment on the heteronomy-autonomy continuum](image)

We can also see the close link between Oscarson’s (1997) three stages of assessment and Boud & Brew’s three self-assessment ‘interests’. The link between the three visualisations of Huttunen, Oscarson and that of Boud and Brew is clear and can also
help us assimilate better the growth and development which Falchikov (2007, p. 135) sees as learning “apprenticeship”, which will be discussed in the section which follows.

3.3.2 Falchikov and apprenticeship learning
Like Boud, Falchikov (2007, p. 138) is interested in assessment from the point of view of its ability to transform and “its potential to enhance learning”. Falchikov views peer-assessment as being very much akin to the traditional forms of apprenticeship where apprentices worked with a master of a particular trade and through his guidance or instruction over a number of years were able to reach a similar level of mastery which was then officially recognised. She reminds us that the apprentices learn not only through observation of the expert, but at the same time learn by watching others who have reached various stages of expertise in their apprenticeships.

In a similar manner, Falchikov (2007, p. 130) feels that peer learning is very beneficial in other forms of education because through the observation of peers of “differing levels of ability and expertise” students are able “to see learning as an incrementally staged process”. Thus, Falchikov (2007, p. 132) is convinced of “the efficacy of peer learning and benefits of one-to-one and small-group learning experiences” and also believes in assessment as a means to provide “rich learning opportunities and social support”.

Falchikov & Goldfinch (2000, p. 288) point out that many instructors may avoid using peer assessment simply because they fear for its “reliability and validity”. By so

---

13 Nunan (2000) reports that he and Benson are also interested in such a notion in the context of language learning and “progress from an apprentice through to a competent, and finally the master practitioner”. Interestingly, Oxford (2003) also refers to the notion of “cognitive apprenticeship” in her proposed model of learner autonomy.
doing, Falchikov & Goldfinch feel that students miss out on ‘learning benefits’, which Falchikov (2007, p. 132) lists as being:

- improvements to cognitive and meta-cognitive competencies
- skills development
- personal and intellectual development
- improved social competencies
- beneficial effects on ‘affective dispositions’, such as confidence.

There are also advantages to be gained by the teacher from peer assessment, which include:

- benefits to the assessment process itself
- teachers gain professional satisfaction from the benefits enjoyed by their students.

Despite the apparent advantages, however, Falchikov (2007, p. 134) is particularly concerned about the effects of grades and marks, not just on students’ scoring, but also on their “timing” and on the “focus of student work effort”. Because of the possible corrupting effects of marking and grades, Falchikov recommends that assessment should involve more than just the awarding of grades. She also suggests there is evidence that grades which have been reached through multiple markers are “more reliable than single marks” and this can easily be accomplished through peer assessment, as was the case with Oral Assessment in the AARP.

Although students may have reservations about their “competence” in being able to award marks “fairly and responsibly”, Falchikov (2007, p. 134) believes that students’ lack of confidence in marking can be overcome by training, and even if the awarding of grades is not involved, training and practice can help to overcome “initial student reluctance” and feelings of inadequacy.

Falchikov & Goldfinch (2000, p. 288) raise two interesting points concerning rating and grading in general. Firstly, they mention that even when there is evidence
of consistency between raters, that this does not necessarily indicate that the grades awarded are fair, this being the case because of the fact that “bias in marking” may be in operation. Not only that, but they point out that teacher marking has also been found to be “problematic” in some studies, so that teacher-generated marks may lack reliability and validity. The first case might be difficult to prove, but instances of the second were found by Matsuno (2007; 2009).

*Why Falchikov supports peer assessment*

Falchikov (2007, p. 135) reminds us that peer assessment is regarded by many as a “means of developing self-assessment skills”. Not only that, but they see it as “a necessary step” in this process. Falchikov, as previously mentioned, strongly believes in the power of “apprenticeship learning”, which she envisages as being able to create “a new autonomous expert” and she also believes in the power of peer assessment procedures which can promote “independence”, “autonomy” and “aid learning in the future”.

Falchikov recommends that to be effective, peer assessment should include the three phases of (1) Modelling; (2) Scaffolding and (3) Fading (see Figure 3.3 below), and through peer support, in what she terms “reciprocal relationships”, learners will experience all the rewards to be had from more traditional apprenticeships. Brew (1999, cited in Falchikov, 2007), sees the “fading of support” as equating with “increasing student autonomy”. Like Brew, Falchikov believes that the responsibilities and interaction derived from peer-assessment processes results in the formation of skills which lead to “the development of enduring assessment competence and lifelong learning” and what she refers to as “learner assessment autonomy” (Falchikov, 2007, p. 140).
Falchikov (2007, p. 139) warns that there is a danger that if peer assessment is conducted in the manner of simply awarding grades, then it is unlikely to impact on students and offer the same learning benefits. Likewise, she warns of the dangers lurking in group peer assessment as there is the inevitable “tension between cooperation and competition”, since schooling is traditionally competitive, so there is the “potential to impair learning”. She feels we should all be aware of the tensions that exist between students and co-workers and the fact that even employers encourage cooperation between co-workers within a company, but encourage competition with rival companies. Thus, learners must be able “to balance these two conflicting modes of working”.

3.4 An overview of research studies most relevant to the AARP

Having looked at assessment from the European perspective and from the global perspective, it is now appropriate to hone in on those studies, predominantly from the Far East, which are closest and of most relevance to the AARP. This overview will begin with studies of Writing Assessment which in all cases involved learner-
assessment being compared with instructor assessment in an HE EFL setting, and will then proceed to similar studies involving learners in Oral Assessment.

3.4.1 Studies of writing skills assessment in HE

While research studies which examine the value of peer review, peer-editing and peer feedback in the development of EFL Writing Skills abound, studies which examine the judgmental or rating abilities of learners, using criteria, either as peer-assessors or self-assessors of writing are very few in number. Moreover, all cases at HE level tend to be concerned with discursive essay writing, rather than the descriptive/narrative genre of the AARP. Another way in which these studies differ is that they do not all involve EFL or ESOL majors; indeed, the studies tend to use participants from a wide variety of disciplines who are attending EFL/EAP courses. We will begin our review with a study conducted in a fairly similar context to the AARP, in the neighbouring country of Bulgaria.

Dimitrova’s (1995) study at the University of Sofia is the only one involving Peer-assessment of Writing with English Philology students, and is also the only one in which the peer assessment task involves placing essays in rank order, according to perceived ability. Twelve 2nd year students were given ten 350-word discursive essays, selected from a total of thirty-nine essays as being representative of a range of abilities, which participants were required to rank from 1 to 10, with 10 as the highest in rank.

The ranking of the 12 students, who assessed the essays in class with a time constraint of 80 minutes, was compared to that of 6 expert raters (Bulgarian EFL instructors from the Foreign Language Centre who were not the course instructors), who took the essays home for a week to assess and rank them.
Since there was considerable variation between student ranking and expert ranking, assessment training was offered, with intervention developing in clear stages, with a particular focus at each stage, making use of both collaborative intra-group and inter-group discussion, before the same peer-assessment was conducted again at the end of each stage. The result was that student assessment of the ten essays gradually came closer to the experts’ assessment, an indication that their initially “different frames-of-reference”, through awareness-raising, shared knowledge of the criteria and training, came closer to coinciding.

Not only did student assessment of the 10 essays improve with time, but a significant improvement in writing ability was also noted. This was attributed to the fact that more channels of feedback, which were less confrontational than traditional ones, had been opened up by these processes.

Training also featured in the study by Saito & Fujita (2004) involving 29 female and 32 male business freshmen, aged 18-19, at a private university in Tokyo, of whom 47 took part in peer-rating and 45 in self-rating, on an English writing course of 90 mins per week for 15 weeks.

Students were assigned to two groups which were taught by two experienced Japanese teachers of English. Instruction was given on the structure of the essay and students were introduced to the rating scales, which used six criteria, which were rated on a Likert scale of 1 to 4. Assessment training, which lasted 90 mins, used sample essays from former students for practice in rating, and for comments in the margins. Students reported back on the scores awarded and the instructor explained why some ratings were preferable to others.

Five copies of a first essay were submitted in the 6th week, with three copies used for peer-assessment by three fellow-students, one copy for self-assessment and
one for instructor assessment. Thus, each student received three peer essays to rate double-blindly, as well as their own to self-assess. After receiving all the ratings and comments from peers and instructors in the 7th week, students submitted a revised version in the next lesson. Instructors rated all the students’ first drafts as well as their revised essays. Students also completed an attitude questionnaire, with 5 items rated on a 5-point scale. The same procedure was followed for the second essay, submitted in the 12th session.

Results from peer-rating of the first essay are not provided, but the researchers’ findings from the second essay revealed a strong similarity between Peer-ratings and Instructor-ratings, a convergence the researchers found interesting given the paucity of training, but there was dissimilarity between Self-ratings and Instructor-ratings and between Peer-ratings and Self-ratings. The attitude questionnaire revealed that feelings towards peers were not adversely affected by the Peer-assessment process.

A two-part study conducted by Matsuno (2007), as part of her doctoral research, shared several common features with that of Saito & Fujita (2004), regarding setting, writing genre, duration of instruction, the Rasch Facets analysis software used, as well as her use of triangulated Self-, Peer- and Instructor assessment. Matsuno’s Pre-Study also shared some common features with the AARP, since (1) only one instructor was involved in assessment, (2) peer- and self-assessment contributed to course grades, and (3) the researcher used qualitative data from questionnaires as well as quantitative data to reach conclusions about assessment success, something which was lacking in her Main Study.

14 Multifaceted Rasch measurement (MFRM) uses modelling to place various factors/facets, such as rater harshness/leniency or task difficulty, on a common scale. The interaction between facets or differences in the functioning of particular raters or items can be revealed.
In the Pre-Study, 25 female and 2 male students, majoring in Chinese were the participants. Students produced their 1st essay in the 8th week and the 2nd in the 9th week, with 5 copies submitted in each case. The criteria evaluation sheet consisted of 16 items, rated on a Likert scale of 1 to 6. In the 9th week students worked through three sample essays in class, using the criteria, and in the 10th week they submitted essays which they had Peer-assessed at home, together with their Self-assessment. The same procedure was followed in the 11th week with the 2nd essay assignment. The teacher assessed all the assignments before looking at the Peer- and Self-assessments.

Statistical analysis revealed that the highest Standard Deviation (SD) was produced by the Teacher-assessor as compared with the Peers, showing that she had produced a much wider range of scores than the Peer-raters. The Peer-raters appeared to be quite lenient, while the Teacher was severe in her rating of both topics. While Self-raters tended to be too strict in their assessments of themselves, Peer-raters tended to be too strict with high-ability students and too lenient with low-ability students, a phenomenon also noted by Cheng & Warren (1999; 2005), while the opposite occurred in the case of the Instructor-rater. Of the 26 students who responded to questions about Peer- and Self-assessment, 7 were unhappy with Self-assessment, while 22 thought that Peer-assessment was a useful learning tool. Some expressed difficulty in using the 16 criteria and in assessing other students’ grammar, when their own was insecure; this may have contributed to lenient marking. Time was also another complaint that was voiced, as were concerns about the weighting allotted to student assessment. Most students felt that only grades awarded by the teacher should be recognised as valid.

In her Main Study, Matsuno (2007; 2009) used 97 student-writers and 91 student-assessors from two different universities and also recruited 3 more instructors
with experience of EFL at university level as anonymous assessors. Students and instructors were trained in rating, in a similar manner, using the same 16 criteria and the same 6-point rating scale as in the Pre-Study, even though MFRM indicated that the rating scale did not work effectively.

During Weeks 1-7, students were instructed in the format, mechanics, organisation and content of essay writing and in the 8th week they submitted 4 copies of a 300-word discursive essay. After training in class with three essays which were evaluated, using the same Evaluation Sheet as in the researcher’s Pre-Study, students were then given 5 Peer essays to assess at home, of which 2 were common to everyone, and while the Researcher rated all the essays, the other three expert-raters, who had no personal involvement in course instruction, were given bundles of 30, 30 and 37 essays respectively to assess at home and these were returned after one month. The expert-assessors received similar training to the students, which was conducted by the researcher herself. As mentioned previously, this time, no qualitative data was collected.

Rasch Facets Analysis produced many misfitting items, including rating criteria, students and experts. Some misfitting students were eliminated while misfitting experts were retained. The Likert scale was condensed from a range of 1 to 6 to a range of 1 to 4 and some misfitting criteria were collapsed. Finally, there were 97 student writers, 81 student raters and 4 expert-assessors, including the researcher. The correlation between the researcher and the other experts was $r = 0.70$ and between the peers and experts was $r = 0.50$. As in Matsuno’s Pre-Study, in the Main Study, Self-assessors were overly severe, due to modesty, while Peer-assessors were overly lenient. Matsuno therefore concluded that teachers should be cautious about including Self- and Peer-ratings in final essay grades.
The disappointing feature in all of the aforementioned studies is that the reasons for conducting the studies and the underlying pedagogical philosophy for involving the learners in assessment processes are rarely made explicit, and therefore the promotion of autonomy does not appear to be the underlying motive. Fortunately, in the case of Oral assessment, the picture, although still a little murky, does appear to be slightly different.

3.4.2 Studies of oral skills assessment in HE
Studies of Peer- and Self-assessment of Oral skills at HE level, which are of relevance to the AARP, will be reviewed in this section, beginning with a study of ESL majors at a university in Hong Kong by Miller & Ng (1996). An interesting, but somewhat unconventional approach to the assessment of Oral skills was taken in this study, involving forty-one 1st year students at City University in Hong Kong.

After a simple 2 hr lecture on how to prepare, administer and grade a speaking test, the students were given one week to prepare a test, with a total duration of 30-40 minutes, to be conducted with another 4-member group from their class. One class teacher sat in on the tests and graded the oral performance of each student, at the same time as the peers, using both letter-grading and numbers as the students were required to do. The study offered each student a quite unique experience as a test-constructor, oral examination interlocutor, assessor and assessee.

In a follow-up questionnaire, seeking the benefits of this experience, students were very positive about their experience in the first two previously-mentioned capacities, and recognised the benefits derived from cooperation with others, but they complained about their own subjectivity in Peer-marking, and the feeling of discomfort in being assessed by their peers. They had mixed feelings about the utilisation of Peer-assessment again in the future, either as assessors or assessees, an
attitude also encountered by Clifford (1999, p. 122), although in the latter case students knew their assessment had been successful. In the case of Miller & Ng (ibid), student opinions were expressed without access to the outcome of the Peer-assessment process, which was actually very positive, with very close correspondence indeed between Teacher-assessment and Peer-assessment, comparable with the AARP Post-Study Oral Intervention Exercise results, in 2009-2010, described in Chapter 5, Section 5.3.4.1.

Patri’s (2002) study was also conducted at City University, Hong Kong, but with 56 native-Chinese students majoring in Computer Mathematics and Information Science, of whom 41 were 1st year students derived from two different English courses, one of which was compulsory and the other voluntary. The class was of 2 hrs duration per week.

The researcher created a research/experimental group (EG) and a control group (CG), the difference with the CG being that they would perform the same tasks without peer feedback. As with the AARP, the researcher was the sole instructor-assessor and each student was required to prepare and give an oral presentation, but unlike the AARP, with topics chosen by the instructor.

Evaluation focused on the four main characteristics of: 1) Organisation and content; 2) Use of language; 3) Manner and 4) Interaction with the audience, but the evaluation form was designed in the manner of a questionnaire rather than a checklist, had 14 items and used a similar 5-point Likert scale, from 1-Poor, through to 5-Excellent. Students were required both to Peer-assess and to Self-assess, using the same criteria as the instructor.

Some practice in presenting was given, with the teacher offering feedback. Training in Peer-assessment was offered with video-taped recordings, but Patri also
made use of live presentations, with a duration of 3 minutes, given by the students themselves, after 30 minutes’ preparation in class.

Each class was divided into sub-groups of 3 or 4 individuals and after each presentation, students discussed it within their groups. It was only after this peer-feedback session that individuals completed their Peer-assessment and Self-assessment forms. In the case of the CG, this feedback stage was omitted. Unlike the training session, no feedback was offered by the researcher to either group.

Conclusions reached from statistical analysis were that peer feedback did seem to make a difference in the Peer-assessment scoring patterns of the EG, bringing them closer to the instructor’s, while in the case of Self-assessment, there were significant differences compared to Instructor-assessment, suggesting that peer-feedback had not greatly influenced S-A scoring. In common with other studies (Matsuno, 2007), both the CG and the EG appeared to overrate low-ability students, though to a lesser extent in the EG.

Unlike the two previously-described studies, the promotion of autonomy was the key aim in the study described by Natri (2007), as is the case in the AARP. Her research was conducted at the University of Jyväskylä, Finland, where she is a French instructor. Although her contact with students of mixed abilities from many faculties is limited to one semester, she believes that it is important to encourage autonomy in their language-learning and support this through assessment processes.

Her intent from the commencement of each course is to raise student awareness of their language abilities in the four skills, through use of the CEFR band descriptors of language competence. Thus, students both identify their starting points in the various skills at the beginning of their course and set their own targets. Natri gathers together the profiles of all the individuals in the group and identifies the mean level in
each of the skills. These mean levels act as a gauge by which individuals can judge themselves, as do the level descriptors.

In the case of speaking skills, students Peer-evaluate various speaking exercises, in addition to presentations, all of which are face-to-face. The grading scheme is on a range of 1-6, from Failed to Excellent and the performance is judged on the five criteria of: 1) sentence complexity; 2) fluency; 3) correctness; 4) pronunciation and 5) the ability to maintain discourse. With time, the Peer-evaluations become more detailed and informative.

The course culminates with Self-assessment of the learning grid created at the beginning of the course and students award themselves a grade for each skill, including Speaking, which counts towards 40% of their final course grade, while 44% is derived from Instructor-assessment and 16% from Peer-assessment; thus, learner-derived assessment plays an important role. Natri accepts that not all students may be satisfied with this system and that no single approach can possibly suit everyone, but still feels it is nevertheless worthwhile aiming for autonomy.

As in the AARP, in Saito’s (2008) study, oral assessment was of a presentation, in this case of 5 mins minimum duration and counting towards 15% of the final speaking-listening course grade.

Like Patri’s (2002) study, the 73 students were divided into experimental (EG) and control groups (CG), with only the EG being given rater-training, which lasted 40 mins and used video recordings of past presentations. Three independent Instructor-raters, who assessed from video recordings, received the same training as the EG.

The Japanese evaluation form, based on Yamashiro (1999, cited in Saito, 2008) and Yamashiro & Johnson (1997, cited in Saito, 2008), used the three categories of: (1) verbal delivery; (2) non-verbal delivery and (3) organisation/purpose, with each
category subdivided into 4 aspects, making 12 criteria in total, rated on a scale reduced from 1-5 to 1-4 points.

Rasch Facets Analysis revealed rating to different frames of reference, with misfitting Peer-raters and misfitting Instructor-raters on various aspects, but none were eliminated. 1st Study results showed EG and CG to be equally close to Instructor-ratings, leading the researcher to believe that misfits could be eliminated through longer training.

In the 2nd Study, 81 1st year students, inexperienced in Peer-assessment, were divided into two EGs and one CG. All groups were taught by the researcher and four independent Instructor-raters assessed video-recorded presentations. All groups received instruction in the 12 presentation skills related to the 12 criterial aspects on the evaluation form, but only the EGs had 5 rater-training sessions of 40 mins focusing on particular aspects, while the CG were given textbook activities related to speaking, listening or grammar. Four independent Instructor-raters had the same training as the EGs, with a duration of 90 mins.

As with rater-training in Patri (2002) and Chen (2006a), after watching video-recordings during which they rated, students discussed their ratings in small groups and then the instructor explained which grades were appropriate and why.

Students presented during weeks 6 to 8, were assessed by their classmates, and in week 9 were given feedback from their peers. Instructors rated from video performances and all raters used the same rating scale.

Data analysis revealed that the longer rater-training helped to remove misfits, particularly among the instructors, who rated more severely than Peer-assessors or Self-assessors. Peer-rating was lenient compared with Instructor-rating and the 1st EG had the most severe raters, while the CG had the most lenient raters; however, the CG
was closer to Instructor-rating on presentation quality. Overall, there were no significant differences between CG and EGs, showing that rater-training had not made a difference. Additional tests revealed this was not due to differences in presentation quality between the two groups.

Of particular relevance to the AARP is Saito’s (2008, p. 576) conclusion that Peer-assessment is “reliable without much training” and that longer training in Peer-assessment does not necessarily bring Peer-rating closer to Instructor-rating, as seemed to be the case in the AARP Post-Study. Saito also questions the usefulness of language assessment skills for non-ESL majors and therefore the time invested in assessment training.

This was not an issue in the case of Chen (2006a), who conducted her research with 23 female and 17 male 2nd Year undergraduate English Majors at a national university in S. Taiwan. Classes, of 15 and 25 students respectively, took a course in English Oral Training, for 2 hrs per week. 13 of the total 40 students had experience of making comments/remarks as feedback in their 1st year, but no experience of rating in Self- and/or Peer-assessment.

The study instruments included an evaluation form, and pre- and post-study questionnaires. The criteria discussed and selected for evaluation were classified and weighted as: Content (30%), Language (30%), Delivery (30%) and Manner (10%). Scoring of the criteria was on a five-level scale, with below 60% considered to be Poor, 60-69% Okay, 70-79% Fair, 80-89% Good and 90% and above considered to be Excellent. The assessment counted towards 5% of students’ final course grade.

The Pre-study Questionnaire sought information about personal background and opinions about peer- and self-assessment, while the Post-study questionnaire
sought opinions about assessment and its effect on learning. Responses were on a Likert scale of 1, Strongly disagree, to 5, Strongly agree.

In Week 1, classes were split into working groups and were informed about the course assessment task of story-telling and about the three stages of (1) Training, (2) Observing performances and (3) Sharing comments and suggestions.

The pre-questionnaire was distributed and, after discussion of key points concerning story-telling, the teacher, together with the students, created a set of criteria for the evaluation form. In Week 2, students watched two video-taped performances, followed by the teacher presenting her assessment and rating of the performances and students also giving feedback on their own evaluations.

After the 2-week training period, each week a group of students was assigned to present stories of their choosing, with teacher and peers assessing each individual performance simultaneously, using the same evaluation form. After the presentations, students discussed and commented on their reflections and opinions within groups and then fed back their opinions to the performers, one group at a time. The evaluation forms were returned after the feedback sessions and the best performance of the bunch was selected by the teacher and the student groups.

A unique feature of Chen’s (2006a) study is that teacher-assessment reliability was tested through a second round of assessment, using recordings of the students’ presentations one month later. Coefficients of overall scorings and of the four assessment categories were 0.87 or over, attesting to reliability.

Comments on evaluation forms revealed participants’ honesty and modesty. Matches in scoring tended to occur between Students and Teacher in cases where the scoring was at the top-end of the scale. Few students overrated themselves, but several rated themselves very poorly, when not rated so either by their peers or their
teacher. There was, therefore, a distinct tendency to underrate, with I-A ranked first, P-A ranked second and S-A ranked last: the opposite of that in the AARP, where Self-assessors generally showed most leniency and the I-R, most harshness. Also, in Chen’s study, the SD for I-A was larger than P-A and S-A, showing that the teacher used a wider range of scores, but Peers were equally discriminating. Pearson correlation coefficients revealed high correlations between P-A and S-A, with 0.87, between S-A and I-A, with 0.87, and between SA and I-A, with 0.70. Chen mentions that agreement tendencies, generally, may have been the result of the feedback and discussion process on performance, which she admits could have had a moderating influence on rating.

Comparisons between the pre- and post-study questionnaires revealed a positive shifting in favour of peer- and self-assessment, with students becoming more positively disposed and confident about it; nevertheless, students suggested reducing the ratio of the grade for student assessment from 5% to 2.4%, due to possible anxiety about assessment accuracy. A t-test revealed students tended to believe peer- and self-assessment to be conducive to learning and improved speaking skills, critical thinking and awareness of strengths and weaknesses, as well as confidence about commenting and evaluating. They did not believe student assessment damaged relationships with their peers.

In another study, Chen (2006b; 2008) researched both peer- and self-assessment of oral production with 18 female and 10 male 1st year undergraduates, of whom 22 were English majors and 6 non-English majors at a national university in S.Taiwan, attending a class in oral training which met for 100 minutes per week. Seven students had previous experience of oral self-assessment, while two had experience of peer-
assessment. The instructor was the researcher herself and, in this study, she was interested in the assessment-autonomy connection.

The evaluation form used was the same as that in Chen (2006a), as was the post-Questionnaire, but this time the assessment program followed six distinct stages, as opposed to three, of (1) Training; (2) Observation; (3) Evaluation; (4) Discussion; (5) Feedback and (6) Response.

As in Chen’s (2006a) previous study, students were divided into groups, informed about the assessment task, and a set of criteria and scoring standards were discussed and established together. At the second meeting, students practised using the criteria with video recordings. Students discussed their impressions and scoring with other group members, the teacher then commented on her own assessment, evaluation and grading, offering feedback and comments on students’ evaluations.

Groups took turns in presenting topics of their choosing for 3-5 mins each. Teacher and peers assessed simultaneously, using the same marking criteria. After their presentation, students discussed their observations together in groups. The peer groups and the teacher provided feedback to the presenters and then the presenters responded to these comments. The presenters used the same evaluation forms to self-assess.

In both assessment cycles, students’ mean scores for S-A were lower than I-A, deviating by five points in the first cycle, but the range of scores was identical. In the second cycle, the mean difference was reduced to 1.4; correlation coefficient values were also stronger. Further statistical analysis revealed evidence of greater matching in the second cycle and a consequent improvement in assessment accuracy.

Chi-square test results of questionnaire comments indicated that student assessors gave most emphasis to manner and delivery, identifying nervousness and
lack of eye contact as the greatest weaknesses, while the teacher-assessor gave more emphasis to content, language and delivery. Generally, the teacher’s comments were more positive and non-judgmental in both cycles.

Chen felt that four points emerged from her study. Firstly, learners not only learned to assess, but learned through the assessment process. Secondly, she believes very strongly in the triangulated processes of peer-, instructor- and self-feedback, for democratic and other reasons. Thirdly, personal and psychological traits clearly, and unavoidably, came into play in the assessment process. Finally, she gives emphasis to the process, rather than the product of assessment. She asserts that while students may not have achieved Littlewood’s (1999) ideal of “proactive” autonomy, through learner-centred assessment processes, they were enabled to practise “reactive” autonomy.

Although not explicitly stated, a similar view emerges from Saito & Fujita’s (2009) study, involving 83 1st year Art students in presenting TV commercials. The aim is to see the effects of group cooperation and joint presentation on peer-rating. A correlation coefficient of 0.74 is produced between peers and the instructor and item hierarchies produced by Rasch Facets Analysis suggest that raters are using a similar ‘frame of reference’. Intra-group reliability in rating is produced in all but one group, where each awarded other members high grades, leading the researchers to conclude that discrimination is possible between individual contributions in most cases, but not all (Saito & Fujita, 2009, p. 159).

Actually, Saito & Fujita make the very interesting point that in implementing learner-centred assessment and in our quest to producer similarity in rating between Peers and Instructors, as teachers we may unknowingly and paradoxically be cancelling out the healthy diversity of opinions which exists between and amongst
peers (Saito & Fujita, 2009, p. 163). Saito & Fujita (ibid) believe that inter-group peer-ratings can be incorporated in students’ grades providing that teachers are prepared to accept “deviations of peer ratings from teacher ratings” (Saito & Fujita, 2009, p. 164). Despite the risks involved, instructors have to take into consideration that if deprived of such opportunities for participation in assessment, students will miss out on developing important learning skills, a view shared by Topping (2003, p. 60) and Orsmond et al. (1996, p. 245; 1997, p. 358), who consider the “developmental process” more important than “accuracy”. Thus, Saito and Fujita (2009, p. 164) believe that “multiple perspectives” are preferable to just the traditional single assessment of the instructor.

Intra-group assessment reliability is also a point of focus in our last study of oral assessment for consideration, in which Peng (2010) used a surrogate instructor to conduct his research on learner assessment of speaking skills in an HE EFL context in N. Taiwan. Through pre- and post-assessment enquiry, he aimed to discover students’ attitudes to peer assessment and to discover any P-A and I-A score correlations and whether these were determined by the students’ level of English.

Cohorts consisted of 43 participants at Upper-intermediate level from various disciplines and 45 equally diverse participants at Lower-intermediate level. Classes met twice a week for 3 hours in total.

A survey questionnaire in Chinese was designed to establish students’ attitudes towards peer-assessment, using a 5-point Likert scale for answers, followed at a later stage by a semi-structured interview for 20% of the 88 participants, who were randomly selected.

Using the criteria from Cheng & Warren’s (2005) study as a starting-point, discussions and negotiation with students helped establish the assessment criteria for
the oral presentations and the designing of a Peer Evaluation and Feedback Form. The 20% weighting, to be awarded to peer-assessment processes, was also decided with the teacher. There were practice opportunities before peer-assessment proper took place.

Of importance is the fact that two kinds of peer-assessment were involved in this study: within-group and group-to-group. While groups were presenting, the other students, who were seated in groups, wrote comments on the Evaluation and Feedback Forms. They then had a few minutes to discuss the performance in their groups and assigned a group grade, the idea behind group discussion being to validate the assessment. Same-group participants also assessed individual contributions using the Within-group Peer Assessment Form.

Student attitudes towards peer-assessment in both 1st and 2nd written surveys were positive and increased in both groups after the activities, although initially, the Lower-Intermediate group was more positive than the other group.

Concerning peer-assessment reliability, as in the case of Saito & Fujita (2009), there were marked differences between group-to-group peer assessment and within-group assessments, leading Peng (2010, p. 101) to conclude that friendship marking had been put into operation, particularly in the case of the Lower-intermediate group.

The researcher found that group-to-group P-As compared well with I-A. While in the first P-A, there was no statistical significance in the Upper-intermediate group ($r=0.51$), there was statistical significance in the case of the Lower-intermediate group ($r=0.97$), while in the second P-A, both groups produced statistical significance compared to the I-A of 0.86 and 0.81 respectively. There seemed to be no differences in assessment on account of language knowledge and, unlike other studies, the I-A was consistently higher than that of group-to-group peers.
3.4.3 Dual assessment of oral and writing skills in HE

It was stated in Section 3.4.1 that studies of peer-assessment of writing, using criteria and rating-scales are relatively rare in the literature; however, studies which combine peer assessment of both writing and speaking with the same groups of students are even rarer.

Cheng & Warren’s (1999; 2005) study stands out as unique in the assessment literature to date as it involves learner-assessment of both speaking and writing in an HE EFL context. The participants consisted of 49 male and 2 female 1st year Electrical Engineering undergraduates at Hong Kong Polytechnic University, where English is the medium of instruction on all courses. The students, none of whom had previous experience of peer-assessment, were taught EAP in groups of 16, 17 and 18, by three class teachers, used to cooperating, working as a team and holding moderation meetings. Classes extended over a period of 14 weeks for 3 hrs per week and involved orientation and training for assessment.

In the first half of the course, students had practice in assessing peers’ spoken and written English, while, in the second half, each class was subdivided into groups of 4-5 students to work on a group project of – a seminar, an oral presentation and a written report, undertaken in that order, and awarded 17 assessment points, based on 1) use of language; 2) peer contribution to the project at various stages and 3) peer-assessment of non-language features of the seminar and oral presentation and the written report. The assessment criteria, used by both peers and instructors, were predetermined by English Department course materials writers, with criteria assessed on a 5-point scale from 1 = Poor to 5 = Excellent.

Group reports were assessed in a similar way, though all the individuals in a group were awarded the same set of marks by the teacher and the members of the other groups. The individual members within a group also assessed the contributions
of the other group members. Interestingly, marks awarded by students counted towards 50% of the total project grade and towards 20% of the total course grade.

A 4-item questionnaire (styled on Burnett & Cavaye, 1980, cited in Cheng & Warren, 2005), distributed to students before and after assessment, monitored their attitudes towards peer-assessment. Three student feedback forms also gauged attitudes towards peer-assessment practices. Students who displayed a marked change in attitude between the pre- and post-questionnaires on three out of the four questions, were invited for an interview; thus, one third of the students were selected for a semi-structured interview which lasted 15 minutes. Students made very revealing statements about their difficulties in assessing the language ability of their peers and in being completely objective and awarding them the marks they truly deserved. Many felt unqualified to judge the English of others and also thought they needed more practice.

Taking into account these concerns, the researchers sought to establish if this insecurity had been conveyed also in the marks they had awarded. Inferences could not be made for report writing, as these assessments were made together by groups of 4 or 5 students; however, from analysis of data, the researchers remained with the impression that with oral assessment, students were not taking into account all the elements that teachers were including and when it came to the assessment of writing, it is quite likely that students may not have been assessing language proficiency at all. Student marking was consistently lower than the teachers’, indicating that the students were using a much narrower range, which means that they had a tendency to mark high-ability peers downwards and low-ability peers upwards, but at the same time their assessment patterns for the non-language criteria were similar.
t-tests were far from consistent across components. For language in the Seminar and the Written Reports, there were significant differences between instructor and students in all groups, while in the Oral Presentation, this was not the case. This may have been due to the elapsing of five weeks between the Seminar and the Oral Presentation, and the possibility of feedback from the Seminars being fed forward to the Oral Presentations. The researchers concluded that P-A grades cannot be used for summative assessment due to unreliability (Cheng & Warren, 1999, p. 312), though they later take a slightly more positive view in their 2005 report, similar to that of Saito & Fujita (2009), that the “long-term, cumulative educational benefits” may be more important than simply trying “to imitate or supplement the assessment behaviour of their teacher” (Cheng & Warren, 2005, p. 112).

3.5 Concluding remarks
This chapter provided an overview of assessment and its relationship to learning, to autonomy in learning and, specifically, to autonomy in language learning. The chapter began with an overview of TEA, moved on to alternative or formative means of assessment, as opposed to purely summative, and outlined the work and research in this direction within Europe and, particularly, in relation to the Council of Europe and the CEFR. The work of the two, most active experts in relation to self-assessment and peer-assessment, who operate outwith the field of EFL, was introduced briefly since their work is of major significance to all subject-areas. The chapter concluded with an overview of all the studies of writing and oral assessment in FL learning, in an HE setting, which are of most relevance to the aims and scope of the AARP.
Chapter 4
AARP Research Methodology

4.0 Introduction
In the previous two chapters, an overview of autonomy and assessment in the field of EFL teaching was provided and the particular significance, relevance and possible inter-relationship between autonomy and assessment at the level of higher education was highlighted, while in the present chapter the aim will be not only to describe how the research on the Assessment for Autonomy Research Project (AARP) was conducted and to outline the main research hypothesis and research questions which were investigated, but also to provide details both of the research methodology and of the pedagogical and assessment methodology and procedures underpinning the AARP.

4.1 The AARP setting
The setting where the AARP was conducted, as mentioned previously in the Introduction to Chapter 1, was the School of English (SOE) of Aristotle University of Thessaloniki, Greece. The SOE is one of four FL Schools within the Faculty of Philosophy, which students enter from high school after achieving satisfactory grades, in specified subjects, in nationwide competitive university entrance exams, known as the Panhellenic exams. The number of places in the SOE is restricted to around 240 annually and preference is given to students with the highest grades, though not necessarily with the highest grades in the English language exam.

The Faculty members of the SOE belong to four different Departments: Theoretical and Applied Linguistics, English Literature, American Literature and Translation & Cultural Studies. In the 1st and 2nd years of their 4-year program, undergraduates are obliged to take some compulsory subjects from each of these Sections and as they progress with their studies, they may choose electives which
permit them to specialise in a particular subject area, such as Translation, Teaching Methodology or American Studies. Many graduates from the SOE continue their studies to post-graduate level either within the SOE or abroad, but the majority of SOE graduates, whether they pursue post-graduate studies or not, become teachers of EFL, either within the state school system or in private FL schools.

What could be considered of importance to the AARP is the fact that SOE graduates are not required, by Greek law, to undergo further training in order to become teachers, and thus the Teaching Methodology courses that they undertake, as well as the way they are taught and tested on other courses, is likely to have a considerable influence on the way they will teach in the future. Thus, the approach to teaching and testing taken on their two obligatory 4-hour language courses, Language Mastery I and II, is likely to have a far-reaching effect, not only on their approach to learning, but on their teaching philosophy and future teaching practices (Sifakis & Sougari, 2010; Sougari, 2011; Sougari & Sifakis, 2007). It was the first of these two courses, LM I, that the AARP was concerned with, during the years 2005-2010.

4.2 Research hypothesis and research questions
The aims of the research were to explore the basic hypothesis which is that there are degrees of autonomy in language learning and that one of the ways in which a greater degree of autonomy can be promoted is through student involvement in assessment of their learning.

In Chapter 2, the two preliminary research questions which concerned: (a) finding a suitable definition of autonomy for the AARP and (b) finding an appropriate model for the AARP, were dealt with. In the case of the definition created, at one and the same time, it satisfied universal beliefs about autonomous language learning as well as being appropriate to local conditions related to autonomy.
and the AARP. In the case of the model designed, it allowed for learner development, along a continuum from heteronomy to autonomy, at the same time allowing for learner differences, for fluctuations in degrees of autonomy, for progression, and even for regression, while allowing, at the same time, for the four areas of responsibility necessary for autonomous language learning, emphasised by Little (1996a, p. 3; 1999a, p. 28) of What?, Why?, How? and With what success?

Having established a suitable definition and model of autonomy for the local conditions of the AARP, the next objective was to establish conditions for the research study, with the aim of ascertaining answers to the four research questions. These four research questions, previously listed in Chapter 1, Section 1.4, but repeated here for convenience, were based on the research hypothesis of promoting autonomy through peer- and self-assessment, and were as follows:

(1) Can learners assess the oral and writing skills of their peers, according to predetermined criteria, with objectivity and reliability?

(2) Using peer-assessment as a stepping-stone, can learners then assess their own oral and writing skills performance with the same objectivity and reliability, using the same predetermined criteria?

(3) Is there evidence that learners can assume ownership of the assessment criteria checklists and exercise judgement in an atmosphere of cooperation and trust?

(4) Is there evidence that the cyclical process of peer- (P-A), self- (S-A) and instructor-assessment (I-A) can combat pre-conceptions about assessment, turning it from a possible constraint to learning into an affordance and thus using it as a learning opportunity which might lead to greater self-direction and autonomy?
The research methods were planned with a view to answering these questions as satisfactorily as possible, using a mixed-methods approach.

4.3 Research design
A literature review was conducted in the year prior to commencing the research, 2004-2005, in order to establish: (a) if any similar studies had been conducted before and (b) if there were any established procedures for this type of research which would be worthy of emulation. Although a few random studies were found related to either speaking (Miller & Ng, 1996; Patri, 2002) or writing (Dimitrova, 1995; Saito & Fujita, 2004), only one study could be found which combined peer- and self-assessment of both speaking and writing (Cheng & Warren, 1999; 2005), but the context of the latter study, reported in two different versions, was altogether different from that of the AARP, being conducted in an EAP context, with Electrical Engineers in a Hong Kong university. Very different research methodologies had been employed in each of these studies, making comparisons difficult and conclusions indeterminable and, most importantly, the main purpose of the studies was not always to use peer-assessment and/or self-assessment as a means to promote autonomy.

4.3.1 The research methods used
Given the complexity of both autonomy and assessment, an attempt to show the relationship between the two necessitated deploying a combination of research methods, so that there is experimental data which aims to show that as learners progress along a continuum of assessment, they are also enabled to progress along a continuum of autonomy (Harris & Bell, 1990), this being the reason why an AARP model was developed (see Chapter 2, Table 2.8).

Thus, the AARP is experimental in nature because there was, first of all, a Preliminary Study to establish the employability of the research instruments and
materials as well as the reliability of assessment results. Once deemed satisfactory, the same methodology was employed in the three-year Main Study, and finally, intervention was deemed necessary in the fifth year of the AARP, which is a Post-Study, to establish if training learners for peer-assessment of writing and speaking could result in changes and possibly improvements in assessment behaviour. The three-part experimental study is attempted through a mixed methods approach to collecting data, involving the collection of both quantitative data and qualitative data.

The research methodology also contains elements of action research since: (a) these are procedures which would have been conducted in class in any case by the instructor, of her own choosing; (b) the research groups used are the instructors’ own groups, which are used as convenience samples; (c) as well as using the assessment data gathered for the generation of course grades, which would be the normal procedure, this same data was subjected to statistical processing and analysis in order to reveal more detailed facts about learner assessment behaviour and to enable comparisons to be made between P-A, S-A and I-A.

Finally, the research methodology, or, more precisely, this ‘pedagogical mix’ of combined teaching, assessment and research methodology, was replicated year-on-year over a five-year period, with students in the same department and with the same materials and classroom conditions maintained as closely as possible, so that the AARP takes on the nature of a long-term case-study, with comparisons possible both within years, between the two research groups in each year, and also across years, between the total ten research groups.
4.3.2 The research instruments used
Many types of instrumentation were used to gather data about the participants in this research study, but in common with the other studies concerned with learner-centred assessment, the two main types of research instrument used in the AARP were:

(1) common criteria checklists for peer-, self- and instructor assessment purposes, for two writing assignments and one oral assignment, used for the duration of the AARP (2005-2010), see Vol. 2, Appendices 6 and 8, for the writing criteria checklists and Appendix 10, for the oral criteria checklist;

(2a) an assessment post-questionnaire used for the duration of the Preliminary and Main Studies (2005-2009), see Vol. 2, Appendix 21 for the Greek version used and Appendix 22, for the English translation;

(2b) a second assessment questionnaire which was specially devised for the Post-Study Intervention (2009-2010), see Vol. 2, Appendix 24.

More details of these instruments will be provided in the sections which follow.

4.3.2.1 Criteria checklists
As previously mentioned, the criteria checklists served a dual or even a triple function in the AARP. They were used as a form of guidance by the students before completing assignments, since they provided a clear breakdown of the criteria by which both written and oral tasks would be assessed by peers, themselves and the instructor alike; thus, the criteria checklists contributed to making transparent the factors by which individual students would both judge and be judged. Each checklist therefore provided indirect guidelines, through the assessment criteria, as to how a task should be completed and, at the same time, direct guidelines, through the criteria and the rating scales, as to how the learners should assess their peers and themselves, when so required.
It should be noted that although it has been common practice for instructors in other subject-areas to negotiate criteria check-lists with the learners, it is not so commonly reported in EFL education. Indeed, in the various published research studies, there is great variability in the number of designated criteria for assessment, in the weighting assigned to particular criteria and in the numerical range used in the rating scale. In the case of the AARP, due to the inexperience of the students and their lack of assessment literacy, it was deemed more appropriate to provide ready-made criteria checklists, designed by the I-R herself, after a number of years of experimentation. The checklists could, of course, be discussed in order to facilitate understanding and dispel any misconceptions.

**Writing criteria checklists**

On the course Language Mastery I (LM I), two separate checklists were provided, which corresponded with the two home writing assignments (see Vol. 2, Appendices 6 and 8). Each of the checklists had been piloted in the previous year and tested for reliability and validity. As previously mentioned, the three copies made of each assignment were accompanied by identical checklists and were used by one peer-assessor, one self-assessor and one instructor-assessor, with the single peer-assessment of writing being conceived of as practice for self-assessment. Each checklist consisted of five holistic criteria, related to interest, relevance, organisation, accuracy and coherence, which were each rated on a scale of 1 – Weak to 5 – Strong. Assessors were encouraged to make corrections, comments and suggestions on the copies of the assignments they had been provided with, and then to use the checklist for rating and calculation of the final grade.
Speaking criteria checklists

A single checklist was distributed to students for speaking assessment purposes (see Vol. 2, Appendix 10), on course LM I, for utilisation in relation to in-class presentations, but its use differed from that of writing assessment in that this time it was used for multiple Peer-assessments, for single Self-assessment and single Instructor-assessment. This meant that, in most cases, students would have used the criteria checklist many times for Peer-assessment purposes before having to use it for Self-assessment, thus eliminating the need for training, and, hopefully, students would have learned to use the checklist with fairness and objectivity.

As in the case of writing assessment, the speaking assessment checklist had five holistic criteria, related to preparation, appropriacy, coherence, cohesion and delivery, which were rated on a scale of 1 - Weak to 5 - Strong. The oral assignment criteria checklist was designed by the I-R herself, following some years of experimentation, and had been piloted in the year previous to the commencement of the AARP, in 2004-2005, and had been tested for validity and reliability. Oral assessments were recorded on individual papers for individual presenters (2005-2008), which were collected in by the I-R, or on specially-designed personal booklets, where as well as space for individual criteria ratings, for each presenter, there was also space for comments on each presentation (2008-2010), thus encouraging students to go one step further and verbally justify the ratings and grades they were awarding.

4.3.2.2 Assessment questionnaires

The assessment questionnaires played a central role in gathering information and data from the participants in the AARP, regarding their attitudes towards and experiences of cooperative, triangulated assessment. The questionnaires were an essential research instrument implemented at the end of each autumn semester, with each of the two AARP research groups, at the conclusion of all the assessment cycles and of the LM I
course. The 25-item questionnaire (see Vol. 2, Appendix 21), in Greek, devised by Antonopoulou (Antonopoulou et al., 2008; Joycey et al., 2010) had been well-piloted and tested for reliability and validity in the years prior to 2005 and so it was used in the AARP Pre-Study and issued to all participants in the study, for completion, at the end of the autumn semester, 2005-2006. It was again found to be reliable and valid and so continued to be used for the 3-year duration of the Main Study, from 2006-2009.

The first page (see Vol. 2, Appendix 22, for the English version) asked for personal information – name, age, place of origin, languages spoken, etc. The remaining questionnaire (see Appendix 22, for the English version) consisted of 22 opinion-seeking questions, concerning the criteria check-lists and peer- and self-assessment processes, which could be rated on a Likert scale of 1 to 5, from Very Much Disagree to Very Much Agree. There were also three further open-ended questions which asked what students thought were the strong points of the assessment method, any problems they encountered, and any recommendations for change which they had to make.

For the purposes of the present study, however, only 10 of the original 22 opinion-seeking questions are taken into account (see Vol. 2, Appendix 23) as these are the questions most relevant to an enquiry into the autonomy-assessment relationship. Also, due to modifications to procedures made in the fifth and final year of the AARP, with Intervention exercises employed by way of training for peer-assessment of the 1st Home Writing Assignment and practice in peer-assessment of the Oral assignment, Antonopoulou’s questionnaire was now deemed obsolete and a new 22-item questionnaire (see Appendix 25), this time in English, was put into place. The advantage of the new questionnaire over the old one was that there were questions specific to assessment of particular Skills, with 13 questions devoted to
Writing intervention and assessment, 7 questions devoted to Oral intervention and assessment, and, finally, 2 general questions on assessment procedures. Again, the questions were answered using a Likert scale of 1 - Strongly Disagree to 5 - Strongly Agree. Another advantage of the new questionnaire was the space provided under each question for students to write comments, if they so wished. This questionnaire had been piloted with another group of students who had undergone similar intervention techniques in the previous semester (spring semester 2008-2009) on the course Language Mastery II, and was found to be reliable.

More information about the way the data derived from use of these instruments was analyzed can be found in this chapter, in Section 4.7, while more information about how these instruments were used, and the results they produced, will be provided in Chapter 5, in Section 5.1.5 for the Pre-Study, Section 5.2.5 for the Main Study and Section 5.3.5 for the Post-Study, with some additional discussion, as and when appropriate, in Chapter 6.

4.4 Background to the AARP
The AARP was conducted over a five-year period and consisted of three very distinct stages (as illustrated in Tables 4.1 and 4.2 below), with (a) a Preliminary Study (2005-2006), the aim of which was to establish if the teaching and assessment procedures followed would yield activities which were useful and results which were reliable, making the procedures worthy of repetition and duplication in the years to follow; (b) a Main Study (2006-2009), which might help opinions to be formed about learner assessment behaviour in general and verify if the outcomes of the Preliminary Study were in any way typical or not, and (c) a Post-Study (2009-2010), in which intervention training techniques in peer assessment of speaking and writing were
employed, enabling any consequent similarities or differences in learner assessment behaviour, as compared with the previous four years, to be identified and noted.

As already stated, the place where the research was conducted, was in the SOE of AUTh and the research groups used were groups from the 1st year English Language courses, LM I, which were assigned to the I-R to teach in the autumn semester of each year. Classes were predominantly female and consisted of mainly native Greek-speakers who were of mixed ability in English, chiefly at B2 level on the CEFR scales, based on Oxford Placement Test (UCLES, 2001) results, and classroom contact with the teacher was 4 hours per week.

4.4.1 The academic calendar and the AARP
Each academic year commenced at the beginning of October and terminated at the end of September, with the autumn semester commencing at the beginning of October and ending in January. The course culminated in a final written exam, with a weighting of 60% of the final course grade, and took place in the January exam session, with a resit session available in September. The AARP is concerned with the procedures and processes which were employed, with the direct involvement of learners, in determining most of the remaining 40% of the final course grade, in relation to speaking and writing skills.

During the first three years of the AARP study (2005-2008), the semester covered a ten-week period, while this was extended, by law, to thirteen weeks of teaching in the last two years (2008-2010). The number of hours assigned to the teaching of Language Mastery I courses, as mentioned above, was 4 hours per week, this being the students’ only language course, in which the four skills of reading, writing, speaking and listening had to be dealt with. Since contact time with students was limited and many entered the SOE with a level of English below what was
desirable, a great deal depended on students being prepared to recognise areas of weakness and being willing to do a large amount of work without direct supervision, at home, in the departmental Resource Centre (RC) or departmental Library, something which was demanding and difficult for students accustomed to a system of spoon-feeding (McKay & Kember, 1997, p. 55). The wide range of abilities within any one group, making it mixed-ability, also meant that those students at the lower end, in terms of language ability, had to work very hard to catch up and reach the appropriate level within a short space of time in order to complete the course successfully.

4.4.2 The assigning of students to groups
Each year the total number of 1st year students was divided alphabetically according to the number of instructors available, which meant that the number of students assigned to each group tended to vary from year to year, but registration was organised electronically and automatically, so that timetables and course instructors for first-year students were predetermined by administrative processes. Initial numbers assigned to groups could be increased by the additional registration of students from previous years and also by the chance inclusion of Erasmus, and, more recently, Erasmus Mundi students who selected particular groups to attend on the Language Mastery I course. It should be noted that a serious factor affecting participation in this study was the fact that attendance on these courses did not become compulsory until the academic year 2009-2010, which meant that participation in peer-assessment and self-assessment processes was, to some extent, voluntary.

In the autumn semester of each year the I-R was assigned to teach two groups of LM I and these are the groups which were used in the research study, which can
therefore be termed convenience samples. The aim of the project was to examine the research questions previously outlined in Section 4.2 and thus (a) look at students’ abilities to engage in peer-assessment and self-assessment of their assignments, as compared with instructor-assessment, and (b) to determine the benefits they might derive from the assignment production and assessment process, including that of exercising a greater degree of critical thinking, reflection, self-determination and, consequently, autonomy. Since there was an adequate number of participants in both the I-R’s groups of LM I to allow comparison, and using the I-R’s groups exclusively for the research would permit much greater consistency in both teaching and assessment practices, it was considered impractical and unnecessary to involve instructors from other Language Mastery I groups or to employ any kind of control group in this particular study.

4.4.3 Factors beyond the I-R’s control

Of course, there were certain features within the research project which were beyond the I-R’s control, such as the size of each sample group and the range of abilities within it, the age and maturity of the students, the degree of commitment and participation in the AARP of individuals and groups, thus making complete homogeneity between groups and years an impossibility. Since participation in assessment activities, as previously stated, was voluntary rather than compulsory, not all students approached the activities with the same relish or with the same degree of sobriety, i.e., there were many factors lying outwith the control of the researcher, apart from motivation, dedication, preconceptions and beliefs, including some interruptions and disruptions to courses and timetables in the early years of the project.
Table 4.1 below shows the number of students who were involved overall in the AARP on the Course LM I and the grey-shading shows how the five-year study was divided into three distinct periods: a Pre-Study, which took place in the year 2005-2006, the Main Study which extended over a three-year period from 2006-2009 and a Post-Study which involved assessment training intervention in the last year of the AARP, in 2009-2010. The number of participants within any one research group therefore falls within a range of 18 to 30 students, making a total of between 41-50 individuals within any one year. For reasons of maintaining the anonymity of the participating groups and for convenience, the I-R’s 10 Research Groups have been assigned names, with the letters A to J, which bear no relation to their actual identity.

Table 4.1  Breakdown of numbers of students involved in the AARP

<table>
<thead>
<tr>
<th>Academic year</th>
<th>LM I groups</th>
<th>n</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-2006</td>
<td>Group A</td>
<td>27</td>
<td>Pre-Study</td>
</tr>
<tr>
<td>2005-2006</td>
<td>Group B</td>
<td>23</td>
<td>50</td>
</tr>
<tr>
<td>2006-2007</td>
<td>Group C</td>
<td>23</td>
<td>Main Study</td>
</tr>
<tr>
<td>2006-2007</td>
<td>Group D</td>
<td>18</td>
<td>41</td>
</tr>
<tr>
<td>2007-2008</td>
<td>Group E</td>
<td>30</td>
<td>Main Study</td>
</tr>
<tr>
<td>2007-2008</td>
<td>Group F</td>
<td>23</td>
<td>53</td>
</tr>
<tr>
<td>2008-2009</td>
<td>Group G</td>
<td>30</td>
<td>Main Study</td>
</tr>
<tr>
<td>2008-2009</td>
<td>Group H</td>
<td>23</td>
<td>53</td>
</tr>
<tr>
<td>2009-2010</td>
<td>Group I</td>
<td>27</td>
<td>Post-Study</td>
</tr>
<tr>
<td>2009-2010</td>
<td>Group J</td>
<td>22</td>
<td>49</td>
</tr>
<tr>
<td>10 groups</td>
<td>10 samples</td>
<td>246 Grand total</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2 below shows both the routine assessment procedures in the AARP, as well as any exceptions which occurred. Basically, the same teaching procedures and assigning of tasks was maintained throughout the five-year period, and the peer-assessment and self-assessment processes on the course LM I involved two writing assignments and one oral assignment. Having established that the teaching and assessment and questionnaire feedback procedures were satisfactory in the Preliminary Study (2005-2006), thereafter the I-R tried in every way possible to
Table 4.2 Research groups and assessment procedures on the AARP

<table>
<thead>
<tr>
<th>ACADEMIC YEAR</th>
<th>AARP – LM I - RESEARCH GROUPS</th>
<th>1ST WRITING ASSIGNMENT</th>
<th>2ND WRITING ASSIGNMENT</th>
<th>ORAL PRESENTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group B</td>
<td>Self-assessment</td>
<td>Self-assessment</td>
<td>Self-assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
</tr>
<tr>
<td>2006-2007</td>
<td>Group C</td>
<td>Single Peer-assessment</td>
<td>*Pre-Self-assessment</td>
<td>Multiple Peer-assessment</td>
</tr>
<tr>
<td></td>
<td>Group D</td>
<td>Self-assessment</td>
<td>Single Peer-assessment</td>
<td>Self-assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
</tr>
<tr>
<td></td>
<td>Group F</td>
<td>Self-assessment</td>
<td>Self-assessment</td>
<td>Self-assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
</tr>
<tr>
<td></td>
<td>Group H</td>
<td>Self-assessment</td>
<td>Self-assessment</td>
<td>Self-assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
</tr>
<tr>
<td>2009-2010</td>
<td>Group I</td>
<td>*Training for assessment</td>
<td>Single Peer-assessment</td>
<td>*Training for assessment</td>
</tr>
<tr>
<td></td>
<td>Group J</td>
<td>Self-assessment</td>
<td>Self-assessment</td>
<td>Multiple Peer-assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Instructor-assessment</td>
<td>Instructor-assessment</td>
<td>Self-assessment</td>
</tr>
</tbody>
</table>

*Deviations from the norm.
follow the same procedures, using the same materials, the same criteria for assessment, setting the same assignments, using the same questionnaires, etc. in order to achieve continuity over the long-term and allow comparisons to be made between research groups in consecutive years. Table 4.2 above outlines the writing and oral assignments for each year and how the assessment for each was conducted, with deviations from the norm indicated by an asterisk. For the reason that assessment processes were considered to be an integral part of the learning process, more information about assessment procedures and assessment cycles will be given, in the sections which follow, together with descriptions of the teaching and learning processes carried out in relation to LM I assignment completion, but it is important at this point to again make clear the distinction between peer-assessment of writing, which involved assessment of a single writing assignment and peer-assessment of speaking, which was achieved through all members of the group each assessing the presenter(s) individually. This distinction is clarified in Figure 4.1 below:

**WRITING SKILLS**

![Diagram of Writing Skills Assessment Cycle]

**ORAL SKILLS**

![Diagram of Oral Skills Assessment Cycle]

*Figure 4.1 Assessment Cycle for Writing Skills and Oral Skills*
Table 4.3  Level of English of SOE 1st Year autumn intake 2005-2009

The above data was provided by Dr Agathopoulou and Ms Groutka of the SOE and is gratefully acknowledged.
This distinction between Peer-assessment of Writing skills and Speaking skills is significant, since in both cases Peer-assessment was being used as the training-ground for Self-assessment and, clearly, much more practice in using the assessment criteria was offered in the case of Oral assessment.

4.4.4 Gathering data about the participants
Since 2003, all first-year students entering the School of English have been subjected to the Oxford Placement Test (UCLES, 2001) at the first or second meeting of their Language Mastery I class and the results have been collated to determine the level of English with which students enter the School. Table 4.3 above shows the approximate general level of English of the majority of 1st year students who entered the SOE 2005-2009, while Table 4.4 below shows the background information for all the participants in the AARP, including their level of English on entry. Thus, if we compare the data in Tables 4.3 and 4.4, it becomes apparent that the level of English in the I-R’s groups was typical of the general trend each year, with a knowledge at CEFR B2 level being predominant in each case.

Unfortunately, due to absences and irregularities in attendance of classes and meetings with the I-R, not all the participants in the AARP took the placement test and not all of them provided information about previous attainment levels in English, but a general overall impression can be gained from Table 4.4, which gives a synopsis of personal data of most of the AARP participants related to gender, age, nationality, level of English according to the Oxford Placement Test (UCLES, 2001), and according to previous public exam certification.
Table 4.4  Background information on AARP participants

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>Group</th>
<th>Participants</th>
<th>Female</th>
<th>Male</th>
<th>17-19</th>
<th>20 and over</th>
<th>Greek/Cyp.</th>
<th>Other nat.</th>
<th>LEVEL AT ENTRY</th>
<th>PREVIOUS LEVEL ATTAINED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C2</td>
<td>C1</td>
</tr>
<tr>
<td>2005-2006</td>
<td>A</td>
<td>27</td>
<td>24</td>
<td>3</td>
<td>21</td>
<td>6</td>
<td>24</td>
<td>3</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2005-2006</td>
<td>B</td>
<td>23</td>
<td>21</td>
<td>2</td>
<td>17</td>
<td>6</td>
<td>21</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>2006-2007</td>
<td>C</td>
<td>23</td>
<td>20</td>
<td>3</td>
<td>20</td>
<td>3</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>2006-2007</td>
<td>D</td>
<td>18</td>
<td>16</td>
<td>2</td>
<td>15</td>
<td>3</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2007-2008</td>
<td>E</td>
<td>30</td>
<td>27</td>
<td>3</td>
<td>20</td>
<td>10</td>
<td>28</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>2007-2008</td>
<td>F</td>
<td>23</td>
<td>19</td>
<td>4</td>
<td>17</td>
<td>6</td>
<td>23</td>
<td>0</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>2008-2009</td>
<td>G</td>
<td>30</td>
<td>25</td>
<td>5</td>
<td>18</td>
<td>12</td>
<td>30</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>2008-2009</td>
<td>H</td>
<td>23</td>
<td>21</td>
<td>2</td>
<td>17</td>
<td>6</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>2009-2010</td>
<td>I</td>
<td>27</td>
<td>23</td>
<td>4</td>
<td>21</td>
<td>6</td>
<td>26</td>
<td>1</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>2009-2010</td>
<td>J</td>
<td>22</td>
<td>17</td>
<td>5</td>
<td>18</td>
<td>4</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td>246</td>
<td>213</td>
<td>33</td>
<td>184</td>
<td>62</td>
<td>234</td>
<td>12</td>
<td>16</td>
<td>72</td>
</tr>
</tbody>
</table>

Levels shown are CEFR levels. U = Unknown
4.4.5 Profile cards and contracts

In order to be able to keep track of the participation of individual learners in the AARP, students were requested to voluntarily complete a Profile Card (see Appendix 3) with their names, place of origin and their hobbies and ambitions and if they so wished, they could attach a photo. The profile cards had a dual purpose, which was to enable the instructor to become more familiar with student names and personal interests and to enable the instructor to keep track of completion of particular tasks and individual progress. No student refused to have a profile card, although this was retained by the I-R, but some chose to leave hobbies or ambitions blank and some chose not to provide a photo, the latter usually through neglect rather than intention.

Another important function of the Profile Card was as a focus and reference point for both the instructor and the learner in discussions about previous levels of attainment in the language and concerning participation in assignments and progress in general. It was therefore an important, albeit a time-consuming, task for the I-R to ensure that the information on the cards was kept up-to-date and that information concerning the triangulation process of self-assessment, peer-assessment and instructor-assessment of both oral and writing tasks was cross-referenced and cross-checked and that the Means of the three were calculated. In this way, students were instantly informed of the closeness of their self-assessment to the assessment of their peers and their instructor and were also kept up-to-speed with the closeness of their own assessment of their peers as compared with their peers’ self-assessment and their peers’ instructor-assessment. By informing individuals about their peers’ assessments and grades as well as their own it was hoped that students would come to treat assessment in a democratic, objective and responsible manner and also move beyond the traditionally competitive and egocentric attitude to grading and evaluation promoted by education systems world-wide (Finch, 2003, p. 64).
The Learner-Teacher Contract (see Appendix 4), which was kept in the possession of the learners, was another important point of reference, the aim of which was to stimulate thinking and initiate a dialogue with the I-R about individual strengths and weaknesses with regard to the skills of Writing and Speaking, and to function as a springboard to action to improve on areas of weakness. The Contract could be used in conjunction with the previously-mentioned Learner Profile card to raise learner awareness and to pinpoint possible areas for development and improvement and identify action which could be taken to overcome possible 2\textsuperscript{nd} language attrition revealed by the learner’s OPT result on entry or from assessment of assignments.

Follow-up meetings could be initiated by the students at any time in the I-R’s office or in the departmental RC, but due to hesitation and shyness, most often the I-R had to initiate and negotiate both the initial and subsequent meetings, and did so with everyone, regardless of language ability. After the initial meeting and discussion, the Contract was signed and dated by both the learner and the I-R, and the learners were then left to assume responsibility and to determine their future goals and courses of action.

4.5 Assessment procedures on Language Mastery I

Students in the I-R’s LM I groups were required to quickly assimilate and take on board two types of assessment with which they were unfamiliar, namely:

(a) Continuous Assessment, which involved ongoing assessment, which was something quite different and separate from the final end-of-semester summative course exam, and consisted of individual home assignments and group in-class assignments, and
(b) Formative Assessment, which was both ongoing and involved student participation in the form of peer-assessment and self-assessment of home assignments.

The significance for the students in terms of the awarding of course grades was that Continuous and Formative Assessment counted towards 40% of the final course grade, while the summative final exam, as previously mentioned in Section 4.3.1, had a weighting of 60%. Completion of assignments and participation in assessment processes therefore played an important role in successful completion of the course.

Students were invited to express any worries or doubts about the assessment process and the instructor, giving examples from her past experience and research on the subject, did her best to allay those fears. The students’ main concerns lay with regard to their relative inexperience in having to judge either their own work or that of others in this way and they aired their concerns that some injustice might be done either to them by others or by them to others. In this respect, to be honest, they were mainly concerned about themselves and others committing injustice by erring on the side of meanness rather than generosity and thus affecting their grades unfavourably.

They felt reassured knowing that in both the case of oral assessment and writing assessment that the work of assessors would be done anonymously, using code numbers. In the case of writing, further anonymity would be achieved by mixing and matching the assignments of the instructor’s two groups of Language Mastery I together, to avoid recognition, perhaps through familiarity with their classmates’ form of handwriting, their code numbers, personal information discussed, or particular styles of writing, etc.
4.5.1 The Assessment Criteria

Concerning assessment, for the simple reason that students, in most cases, were being placed in a situation where they would assess others and themselves for the very first time, and were being put in a position of informed assessment where responsibility for assessment and grading is shared equally between Self, Peer and Instructor, they were presented, in each instance, with a set of criteria:

(a) which were already specified and produced by the I-R, without negotiation taking place amongst learners or between learners and instructor to define them;¹⁵

(b) which were in the form of a simple check-list with five criteria on a gradation scale from 1-5, which did not involve complex operations to use or complex mathematical formulae to calculate;

(c) which were known to them in the pre-production stage of the assignment, and

(d) which became more familiar to them with time and use through the introductory peer-assessment process, particularly in the case of oral skills, since it was what Chen (2008, p. 244) refers to as an “iterative” process.

The criteria selected for the assessment of assignments in the case of both writing and speaking tasks were consciously holistic in nature and avoided very specific categories such as ‘grammar’ and ‘vocabulary’ chosen for inclusion by many other researchers. The reason for avoiding such category labels in the criteria checklists was that given the mixed ability of the groups, some students might be aware of their own superiority or inadequacy in these areas and consequently feel either over-qualified or under-qualified to judge their peers in this respect, particularly after discussions with the instructor about their learning contracts, when they were

¹⁵ Until recently, researchers such as Boud (1989), Little (2002) and others had insisted that students should be involved in negotiation of the criteria themselves or even work with their own criteria exclusively, but this has recently been negated by Orsmond et al. (2000) who found that even with student negotiation of criteria, ownership of the criteria is not guaranteed.
encouraged to try and identify areas of strength and weakness, so had already reflected on their language abilities and inadequacies.

On the matter of negotiation of the criteria checklists, many experts in the field recommend a process of negotiation in order that the raters using the scoring criteria will feel greater ownership of them and have created them after a great deal of thinking and deliberation, but the I-R’s experience in previous years in this respect found that:

(a) negotiation of criteria involved an extremely time-consuming process, and

(b) lack of expertise in the subject-matter to be assessed made it very challenging and difficult for learners to create a balanced set of criteria.

The I-R therefore preferred to save time and effort in the run-up to the more tricky business of actually applying the criteria, and offer ready-made criteria which were fairly simple in application and had proved reliable in the past. Besides, in their more recent work, Orsmond, Merry & Reiling (2000, p. 33) have suggested that students may actually be forced to think more deeply about criteria that are ready-made and given and that even when criteria are co-constructed with the instructor, it does not mean that they will be interpreted by all participants in the same way (Orsmond et al., 1997, p. 365).

In writing assignment assessment and oral assignment assessment, both when presented with the assessment criteria for the first time and when the criteria were put into use for the first time, the instructor checked with the students that they fully understood the criteria and also felt confident about the assessment procedure as a whole, so that more than enough opportunities were given for students to ask questions or clarify doubts concerning using the criteria.
4.5.2 The Assessment Cycle

The I-R imposed a strict cycle of assessment on the learners so that after submission of the Writing Assignments, the triplicate copies were divided into three piles designated for Peer-assessment, Self-assessment and Instructor Assessment and this was the order in which assessment was conducted. After what was deemed a satisfactory passing of time after submission, in class-time learners were given the assignment of a peer to read, correct, comment on and evaluate using the specified assessment criteria. This was conducted ‘blind’ in that it was assumed that students could not identify the writer and therefore show bias in their assessment. They could take as long as they wanted to complete the assessment and could ask questions of the instructor if they had any. Boud (1995, p. 202) emphasises the fact that there should not be restraints on time and that students should be given sufficient time to enable them to construe a “sufficient appreciation” of their classmates’ work and to enable them to construct “valid comments” on that work. If the process is in any sense pressured or rushed, there is the danger of impressions formed and comments given being too superficial (Boud, ibid).

After adding their own code number, and calculating the final grade, they would submit their assessment to the I-R. Students were encouraged to work on their own on this task without consulting or chatting with neighbouring students, which could be distracting not only for their neighbours, but for others. Later on, not only did the Instructor-Researcher record the grades awarded to the Writer on their personal Profile Card, and by whom, always using codes to maintain anonymity, but also recorded the grade awarded and to whom on the Profile Card of the Peer-assessor.

---

16 The conducting of peer- and self- assessments in class-time occurred in the Main Study and Post-study periods. Prior to this, these assessments took place at home so as not to infringe on valuable and limited class contact time, but students were not always diligent about, or timely in, returning these and some were even lost.
Figure 4.2 The LM I Writing Assessment Cycle, with feedback

The I-R also recorded the scoring of individual criteria\(^{17}\) as well as the overall Peer grade total awarded to each individual writer electronically, and stored these for future research. Figure 4.2, above, shows the general assessment cycle and the order in which the triangulated assessments took place, followed by feedback from the triangulation process. Again, after the passing of what seemed like a reasonable period of time, which Boud (1995, p. 202) claims enables learners to sufficiently “distance” themselves from their work, the I-R distributed copies of the 1st Home Assignments designated for Self-assessment and made sure that learners found themselves confronted with their own assignments for assessment and requested that they treat them with the same degree of objectivity with which they had supposedly confronted the work of their peers (Black & Jones, 2006, p. 8). Clearly, students found this an altogether more challenging exercise to deal with, but did their best to do as requested. Most obvious, was the lack of comments and corrections to their own

\(^{17}\) The recording of the scoring of individual criteria electronically is a lengthy and time-consuming business and so was not always completed between 2006-2008, although the raw data with true scores has been stored. Very few researchers stress the significance of the scoring of individual criteria, but clearly there is considerable scope here for future research.
work, and sometimes the obstinate refusal to find fault with it at all, with students emphasising the fact that they would not have submitted their assignments in the first place if they had not believed they were perfect, an attitude encountered by other researchers (Hanrahan & Isaacs, 2001, p. 59). The completed assessments were gathered in by the I-R and the grading and assessment was recorded in the same way on the Profile Cards and on the computer.

It should be noted that students who for one reason or another were present in class but had not completed a Home Assignment, were not able to take part in these processes until such times as their own assignment had been submitted and they were therefore given an alternative activity to complete during assessment activities.

In the final part of the assessment cycle the I-R took the pile of assignments designated for assessment by her and using the same criteria, and in the same atmosphere of anonymity, and in a place where there would be no interruptions or distractions, the I-R assessed the 1st Home Assignments. On completion of the assessment process, the I-R recorded her grades, in the same way, on the Profile Cards and electronically, so that assessment records for the Home Assignment now only needed the average of the three Self-, Peer- and Instructor Scores in order to complete the triangulated assessment process. The three totals were added together using a mechanical calculator and the average of the three grades recorded on the Profile Cards.

In the final phase of the assessment cycle, students were invited for consultation in pairs with the instructor, and were given the three copies of the corrected assignments to peruse, and in some cases were provided with a photocopy of the instructor’s corrections. They could therefore see the closeness of their self-assessment to that of their peer and the I-R and were also informed of the closeness of their peer-assessment with that of the I-R and the peer-writer’s own self-assessment.
Thus, the feedback they received about their own assignment, differed significantly from conventional feedback, where students pay most attention to the grade they receive, rather than any comments or corrections, since in this case, the feedback was derived from more than one channel. It also differed due to the fact that the feedback they received was not just concerned with their achievements in their own assignment, but was also related to their success in their assessment of their peer’s assignment, in which assessment was also triangulated.

4.6 Pedagogical procedures for writing and oral production

The main aims of Language Mastery I are the oral and written production of descriptive and narrative discourse. With this view in mind, a number of speaking and writing tasks are completed in groups in class time, and two writing assignments and one oral assignment are prepared at home. As can be seen from the breakdown of participants’ language ability on entry to LMI (see Table 4.3 for the general 1st Year breakdown and Table 4.4 for the AARP breakdown), there is a wide spectrum of ability amongst students and therefore the homework task set must be such as to enable students to accomplish the task with their varying degrees of language sophistication. There are also some elements of what Fonseka (2003, p. 147) refers to as the “carnivalesque” in the materials, on which the 1st Home Writing Assignment is based, in order to capture the interest and imagination of the learners.

4.6.1 Procedures for the 1st Home Writing Assignment

The 1st set home writing assignment (see Appendix 5a), which came early in the course, was related to describing people and consisted of two altogether separate paragraphs which were produced by the students based on two quite different teaching and learning procedures and processes. The first paragraph was a hypothetical one based on:
(a) a pre-listening gap-filling exercise related to the lyrics of the Suzanne Vegas song ‘Luka’ where all possible answers were vetted for suitability and discussed (see Appendices 5b & 5c);

(b) a listening exercise in which students checked which of all the possible answers discussed were the ones actually contained in the lyrics (see answer key in Appendix 5d);

(c) a note-making exercise in which students were encouraged to make some guesses about Luka’s character and the nature of his/her problems, before attempting to write the paragraph (see Appendix 5b).

The song is open to many different interpretations, starting with the gender of Luka, which makes it challenging and interesting for students, as there are no ‘correct’ interpretations and the nature of Luka’s character and the circumstances in which he/she lives are open to hypothesis.

The second paragraph was a short description of the class-mate that students interviewed using a short ‘fun’ questionnaire (see Appendix 5e), taken from the Pilot Edition of Swan’s Cambridge English Course 3 (which was piloted in the former Direct Teaching Operation of the British Council, Thessaloniki and did not appear in the final edition published by C.U.P), but was used after seeking the author’s permission. It is the kind of questionnaire which appears in popular magazines of the analytical socio-psychological type, which most people enjoy completing. Students were asked to answer the questions themselves using the Likert-type scale of 1 to 5, from ‘Yes, definitely.’ to ‘Definitely not!’ and then to question their partners and record their answers, in order to prompt oral communication and arouse interest in a class-mate. With the two sets of answers to the questions, any great similarities or differences in character became apparent and we discussed these student findings in class together as an ice-breaking activity, but also to create a stimulus for writing.
The objective in the paragraph-writing in this part of the assignment was to encourage students to link related ideas together and to select from among all the information they had gathered about their interviewee what they considered to be the most salient and interesting points and to connect them together following the structure of a paragraph, which paves the way for structure in future essay-writing, of a Beginning - Middle - End.

Before completing this assignment, students were given practice in linking sentences and linking related and contradictory ideas together as well as practice in varying sentence structure, in order to avoid repetitious monotony, using exercises from Stephens – Practise Advanced Writing, as well as Pincas, Hadfield & Hadfield – Writing in English, 3 and Tsoukala-Smyrni & Wildig – Write Away, Book 1.

Students were requested to submit their two-part assignments (see Appendix 5a) in triplicate copies, with just their code-numbers, by a specified deadline. By providing learners with the assessment criteria (see Appendix 6) before they began the writing process, they were aware of the criteria by which they would both judge and be judged for this particular assignment. This process ensured that there was no “hidden curriculum” with regard to grading schemes (Bergenhenegouwen, 1987; Sambell & McDowell, 1998) and brought everything out in the open. It also ensured that the assignments were completed in a climate of “informed awareness” (Shepard, 1992; 2000) and helped in making the process of assessment more democratic by reducing the ‘power divide’ between instructor and learners, since learners would be trusted to award ‘real’ grades to both their peers and themselves on an equal footing with the instructor, something which MacAlpine (1999, p. 16), Orsmond et al. (1997, p. 358) and Stickler et al. (1999, p. 286) consider is essential if learners are to take the whole process seriously.
In each year of the AARP, from 2005-2009, the same procedures were followed with the 1st Home Writing assignment. What follows is a description of the difference in procedures followed in the Post-Study, 2009-2010.

4.6.1.1 AARP Post-Study 1st Home Writing Assignment Intervention Exercise
Given that the assessment results for the 1st Home Writing assignment in the AARP Main Study, which will be presented in Chapter 5 and discussed in Chapter 6, were variable and differed in many aspects from those of the AARP Pre-Study, it was deemed appropriate to undertake a form of Intervention in the Post-Study. This decision was influenced by the fact that a great deal of emphasis, in the related literature, is placed on the importance of training learners for Peer-Assessment. In the first four years of the AARP (2005-2009), learners had not undergone training in peer-assessment for the 1st Writing assignment, but rather, it was taken for granted that students were capable of peer-assessment and the processes of peer-assessment were, rather, used as training for self-assessment and were thus utilised as a stepping-stone to Self-assessment, both in the assessment cycle of writing skills and of oral skills. The I-R decided to implement some intervention techniques with her groups, albeit on a very limited scale, to see if some kind of training for peer-assessment could make a difference to assessment behaviour and bring its accuracy closer to that of the instructor. These interventions were embedded in the standard procedures for production of the 1st Home Writing Assignment, described previously in Section 4.6.1, and of the Oral Presentations.

In the Post-Study IE, the instructor followed the same teaching procedures and the same preparation for the 1st Home Writing Assignment as in previous years. A deadline for submission was set in the usual way and students were asked to submit three copies of the assignment anonymously, using a code number given by the instructor. They were also introduced to the assessment criteria that would be used in
the triangulated assessment process, so that they could produce their assignments in full awareness of the criteria with which their writing would be judged.

By way of Intervention, about a week after the assignments had been submitted, the I-R selected three assignments from previous years at random (see Appendices 11 a, b & c), and having put them together with the assessment criteria (see Appendix 6) and an assessment sheet (see Appendix 12a), the instructor requested that the students take the assignments home and bring the completed assessments with them to our next meeting. They were also requested to come with ready comments about each of the three assignments. These comments would form the basis of a discussion about the assignments in class. Their scoring of the assignments would be recorded on the assessment sheet which they would submit to the I-R.

At the next lesson, the students were given a paper on which the authentic assessment of the Self, Peer and Instructor of each particular assignment in the year it was submitted was recorded, with space left for the present-day assessors to add their scoring to permit easy comparison (see Appendix 13a). After their scoring had been copied from the assessment sheets to the new photocopies, the I-R collected in the homework assessment sheets for later scrutiny and analysis. A discussion then ensued about the sample assignments and students justified their opinions and the grades they had awarded. Students were then given two more random examples to peer-assess and grade, in class this time (see Appendices 11d and 11e), for which they recorded grades (see Appendix 12b) and they were then given a ‘real’ sample from a classmate which they were asked to peer-assess in the same way. The hope was that they would use the same amount of objectivity with their classmate’s assignment as they would with the ‘mock’ assignments. Students were then given the original Self-, Peer- and Instructor-Assessment grades for the last two of the five ‘mock’ assignment
assessments (see Appendix 13b), which again, after discussion in class, they were asked to look at more closely at home.

The results of the 1st Writing Assignment IE in the Post-Study can be viewed in Chapter 5, Section 5.3.2.1 and the results of the 1st Writing Assignment assessment in both ‘tainted’ and ‘pure’ forms can be viewed in Section 5.3.2.2. An anonymous example of how one student performed in the IE can be found in Appendices 14a & 14b. The high degree of coincidence in rating of the 5 samples, between student and the I-R, is apparent, while the deviation in rating of the same student’s real peer assessment as compared with the I-R, is striking.

4.6.2 Procedures for the 2nd Home Writing Assignment

By the time learners were asked to produce the second home writing assignment, they were approximately half-way through the course and were now familiar with course objectives and the standard of writing expected of them in relation to description and narration.

Following on from focus in classwork on both the physical description of people and description of their character, and exercises on essay writing and construction taken from Stephens – Practise Advanced Writing and Tsoukala-Smyrni & Wildig – Write Away, Book 1, students were presented with the topic of their next Home Writing Assignment, which again consisted of two parts (see Appendix 7), a descriptive essay of the person they either most admire or the person they would most like to be, and also a paragraph narrating a first meeting with a friend. In the case of the essay, the person chosen could be either living or dead, famous or not, from any walk of life, and the paragraph could be true and from any part of their life till now or could be derived from their imagination. In the case of the essay, new assessment criteria were introduced, related to interest, relevance, ideas and linkage, accuracy and construction (see Appendix 8), while in the case of the paragraph, the previous
familiar criteria were given, related to conciseness, relevance, organization, accuracy & appropriacy and interest (see Appendix 8). A deadline for submission was again given, trying as much as possible not to interfere with deadlines for assignments given by other course instructors. The criteria were also made known to class members well in advance of submission.

Similar procedures were followed for submission of the 2nd Writing Assignment as with the 1st Writing Assignment, and thus peer-assessment, self-assessment and instructor-assessment and the results of assessment were recorded in the same way and once the process of assessment triangulation was completed, learners were invited to come and receive feedback on their own assignment and that of their peer.

4.6.2.1 AARP Main Study 2nd Writing Assignment Variation (2006-2007)
Although all the procedures for the 2nd Home Writing Assignment and its assessment, as described in the previous section, apply to the Preliminary, Main and Post-Studies, it would be remiss not to mention a slight variation in procedures, but nevertheless a significant one, which took place in the second year of the AARP, in 2006-2007, the results of which are presented in Chapter 5, Section 5.2.3.1. The variation concerns assessment procedures in relation to Self-assessment of the 2nd Home Writing Assignment, in the 1st year of the Main Study, and it revealed some interesting and surprising results.

Following the relative success in assessment of Writing Skills in the Pre-Study, the I-R believed that it would be interesting to try and establish how the critical thinking skills of the learners are affected by peer-assessment processes and how their self-assessment judgment is affected by reading the assignment of another student and correcting and assessing it. After having discussed the assessment criteria checklist for the 2nd Writing Assignment with the students and provided them with copies (see Appendix 8), when setting the assignment as homework, in this instance,
the I-R informed the students that they would be required to assess their own assignments twice i.e. together with their assignment they would submit a completed self-assessment sheet, and some time after dealing with peer-assessment, they would be requested to assess their own assignment a second time in class.

While 20 of the total 22 participants from Group C conceded to this request, only 13 of the total 18 participants from Group D did so, and since these assessment processes were to some extent voluntary, nothing could be done about this. The results for this variation in procedures, which are rather unexpected, can be found in Chapter 5, Section 5.2.3.1 and are further discussed in Chapter 6, Section 6.1.3.1.

4.6.2.2 AARP Post-Study 2nd Writing Assignment Non-Intervention
Due to lack of time, in the Post-Study, in the year 2009-2010, although Intervention exercises (IE)s had been employed in the case of the 1st Writing Assignment, the I-R took the decision not to employ IEs in preparation for assessment of the 2nd Writing Assignment. Even without a second set of writing IEs, however, some quite significant changes were noted in the learners’ assessment behaviour between the 1st and 2nd assignments, and for this reason two sets of results, one with and one without the non-participants in the 1st Writing Assignment IE were produced for the assessment of the 2nd Writing Assignment. These assessment results can be seen in Chapter 5, Section 5.3.3.

4.6.3 Procedures for the Oral Presentations and Assessment
Oral abilities on the Language Mastery I course were assessed by means of an oral presentation which was conducted by each student before their peers, either in class, forming part of the lesson, which they were encouraged to do, or on a special day assigned to their whole-class group for presentations and their assessment. Since 10% of the total course grade was assigned to speaking, the preparation of the presentation
and its delivery were of quite considerable importance. Each student was required to speak for 3 minutes and their total presentation time was not to exceed 5 minutes. Students could opt to present alone, in pairs, or in small groups, with the same amount of time allotted to each individual.

A great deal of freedom was given to students concerning their choice of topic, related to narrative and description, with just a list of possible suggestions given (see Appendix 9) and students were given no particular guidelines or training, but were encouraged to consult the I-R during office hours if and when they had problems choosing a subject or how to present it, which they often did. Students were encouraged to speak in class through various pair and group-work activities using English as the means of communication, but the presentation was the only formal occasion on which individual students were required to address the whole class. Having said that, given the fact that attendance in class was not compulsory on this course during the academic years 2005-2009, the size of the audience could vary quite considerably from one occasion to another when presentations were executed.

For assessment purposes, students were given separate sets of assessment criteria to complete for each individual, both for peer-assessment and self-assessment purposes (see Appendix 10). The instructor therefore depended on individual students to identify the presenter, by code or name, and themselves as the assessor, by code or name, on each slip of paper containing the criteria checklists. Bundles of peer-assessment papers were collected after each individual presentation, in preparation for recording and calculation, together with the self-assessment, completed immediately after the presentation, and filed together with the instructor assessment papers.

In an attempt to simplify administration of the Oral assessment process, in the last two years of the AARP, a change was made, from using individual criteria checklists for each participant in each presentation, to using assessment booklets,
which contained the names and student numbers of each participant, as well as space for the five assessment criteria grades, the addition of those grades and the final calculation out of 10. Space was also provided alongside these boxes for comments on the presenter’s performance, by way of justification for the grade awarded. A single criteria checklist, which remained blank, except for the assessor’s student number, formed a cover to the booklet. This maintained anonymity and meant that the booklet could be reused by the assessor many times over. The I-R collected in the booklets after each stint of assessment and recorded the grades awarded in the usual way.

Given that no video recording facilities were available, and a second viewing of presentations was not possible, it was important that there be no disturbances or distractions in the room and that students and the instructor should be totally focused on listening and assessing.

As well as using the scoring criteria, and recording her grades, the I-R also made notes of both good points and weaknesses in each presentation in case presenters wished to have more detailed feedback concerning their oral performances at some further point in time.

At a later date, the instructor recorded the self-assessment and the instructor assessment grades on the profile cards and aggregated the peer assessment scores and worked out the mean peer score. The self, peer and instructor scores were then aggregated and divided by 3 to calculate the final grade for the Oral Presentation out of 10%. These grades were also recorded electronically for future reference. The assessment papers were also consulted once again and the actual scoring of individual criteria by each assessor, peer-, self- and instructor, for each presenter was also recorded for closer analysis at a later date.
4.6.3.1 AARP Post-Study Oral Assignment Intervention

A reading of Chapters 5 and 6, which follow, reveals that during the first four years of the AARP, 2005-2009, with the exception of only one of the eight research samples, which was Group E in the Main study, the assessment of Oral skills had produced fairly consistent and satisfactory results. Nevertheless, influenced by reports in the EFL assessment literature of the seeming judiciousness of preparation and training of learners for Peer-assessment, the I-R decided to implement Oral Intervention Exercises (IE)s in the Post-Study in order to examine any changes or improvements that practice in using the assessment criteria in mock conditions might bring to the use of assessment criteria in real conditions.

Since no video recordings of Oral presentations from previous years were available for the purpose of the IEs, the I-R set out to find past LM I students, from years other than 1st year, who had some experience of presenting and who for a small reward would be willing to act as guinea-pigs for peer assessment practice purposes. The volunteers were required to simply come into our class at a specified time and to execute their presentation on any suitable subject of their choosing for the required length of time. They could also choose whether to present alone or with a friend. The presenters were informed that 1st year students would be assessing them, but the presenters were not in any way involved in the assessment process, nor was any kind of discussion about their performance entered into in their presence. The volunteers were used as assessment guinea-pigs, pure and simple, but were useful to the assessors, in that they also provided a demonstration of how a presentation could be prepared and performed, as well as providing a rehearsal for how oral assessment was conducted.

For the purposes of this Oral IE, Groups I and J had to be put together since it was not possible to have the presenters deliver their talks twice over. Apart from the
room being a little more crowded than usual, this did not cause any particular problem, since each student would only be involved in oral peer-assessment, using the oral criteria checklists which they had already seen and had previously had the opportunity to discuss with the I-R. On the day of the presentations, the criteria check-list formed a cover to maintain privacy in marking (see Appendix 10), and an assessment sheet with spaces for presenters’ names and the criteria grades, the total for the criteria grades and the final grade out of 10 was provided below it, against each presenter’s name (see Appendix 15). Space was also provided for comments.

In some cases, assessors arrived after the presentations had begun and so they began to assess from the commencement of the next presentation. There were nine presentations in all, on a wide variety of subjects, delivered in a wide variety of ways. When all the presenters had left the room and the assessment sheets had been collected from students, with their student numbers attached, discussion then ensued about the advantages and disadvantages of each presentation and in terms of assessment what each presentation was worth and why. The I-R collated and recorded all the assessment sheets and the sheets, once processed, were then returned to the students together with a sheet on which the I-R had recorded her assessment and grading for each of the 9 presenters, to permit comparison (see Appendix 16, for an example).

The results of the Oral Assignment IE can be seen in Chapter 5, Section 5.3.4.1.

4.7 Analysis of the Assessment Data
The research instruments employed in the AARP, namely the criteria checklists applied to writing and speaking assignments, and the questionnaires deployed on completion of the assessment processes, and the main aims and purposes of each, have already been described in this chapter, in Section 4.3.2. In the sections which
follow, what will be described are the ways in which the data derived from use of these two research instruments was processed and analysed.

4.7.1 **Analysis of data gathered from use of criterial checklists**

As was previously explained in Section 4.3.2, the criteria checklists used in assessment of writing and speaking skills were one of the two main research instruments used to gather data concerning assessment processes on the AARP. There was no alteration to either the two criteria checklists used in assessment of the two writing assignments (see Appendices 6 and 8) or the one criteria checklist used in assessment of oral assignments (see Appendix 10) for the 5-year duration of the AARP.

In order to gain as complete a picture as possible regarding the outcome of the triangulated processes of assessment, applied to both speaking and writing skills in the AARP, and due to lack of consistency in methods of analysing assessment data in existing studies of assessment, it was deemed necessary to subject the statistical data to three different analytical processes, which are outlined below.

All of the data gathered concerning Self-assessment (S-A), Peer-assessment (P-A) and Instructor-assessment (I-A), from each assignment completed on LM I, in each year of the AARP was subjected to the same methods of statistical analysis. The Means derived in each case, from each type of assessment, whether S-A, P-A or I-A, for each group, for each assignment, for each participant, were subjected to One-Way Analysis of Variance (ANOVA), with a probability reference value of $p = 0.05$. When there were no significant differences, this was an indication of alignment in assessment between S-A, P-A and I-A. In cases where significant differences were found, further analysis using the Tukey-Kramer Comparison Test of Pairwise Mean Differences was deployed in order to gain a better understanding of the relationship between S-A, P-A and I-A in these instances.
Secondly, in order to gain another perspective on assessment outcomes, assessment data was also analyzed using the Pearson product-moment correlation coefficient ($r$) to determine possible relationships (degree of positive or negative agreement) between S-A and P-A (S-P), S-A and I-A (S-I), P-A and I-A (P-I), using a probability reference value of correlation significance, $p = 0.05$.

Thirdly, Chi-Square Tests were also conducted, with the pooled data from the two research groups each year, to determine any Self:Instructor (S:I), Peer:Instructor (P:I), as well as Self:Peer (S:P) deviations in rating frequency ratio patterns, year on year (see Vol. 2, Appendices 18, 19 & 20). Further explanation about the purpose of these tests can be found in Chapter 5, Section 4, where details of the rating frequency ratio patterns are presented.

The same statistical procedures were followed in the Pre-Study (2005-2006) with Groups A and B, in the Main Study (2006-2009), with Groups C, D, E, F, G and H and in the Post-Study (2009-2010), with Groups I and J. Details of the findings from statistical analysis of the data in each of the stages of the AARP, concerning assessment of LM I writing and speaking assignments, are outlined in detail in Chapter 5.

4.7.2 Analysis of data gathered from use of assessment questionnaires

Details of the questionnaire used in the Pre- and Main AARP Studies have already been described (see Appendices 21 for the Greek version, Appendix 22 for the English version and Appendix 23 for the abridged AARP version) in this chapter, in Section 4.3.2. Questionnaires were distributed at the end of each autumn semester in order to gain greater insight into learners’ beliefs and perceptions about the assessment process, thus providing very useful qualitative data to supplement the quantitative data derived from use of the criteria checklists.
However, in order to give substance to participant responses in the assessment questionnaires, which were, for the most part rated on a Likert scale of 1 to 5, this ranking data was treated quantitatively. As in the case of analysis of data derived from criterial checklists, the reason for these two different forms of analysis and presentation in the case of the assessment questionnaires was, firstly, in order to try and obtain analyses of the data from more than one perspective, and, secondly, in order to obtain a more detailed and complete picture of student perceptions than had been offered in previous studies.

First of all, frequency counts were conducted on the responses to each question by each participant and the data was transferred to Excel documents to await further processing. The data was then processed in two different ways:

(1) using percentages which are displayed in the xyz graphs, for the ten questions relevant to the AAPR, in Appendices 26a to 26j for the Pre-Study, and Appendices 27a to 27j for the Main Study. These graphs permit quick visual comparisons to be made between groups, but verbal commentaries are also provided by the I-R and random comments from participants related to each questionnaire item, also accompany the graphs, providing further elucidation concerning feelings towards and perceptions of assessment processes.

(2) according to the means of responses for each year, which were subjected to $t$-analysis, the results of which appear in Table 5.5 for the Pre-Study and Tables 5.16, 5.17 and 5.18 for the Main Study, in Chapter 5. This form of analysis again permits easy comparison of attitudes towards assessment on the course LM I, between groups of student in the same year, and between groups of students in different years of the AARP.
The Post-Study questionnaires were also collected at the end of the autumn 2009-2010 semester and the data was saved and processed in the same way, this time for all 22 questions. The relevant $xyz$ graphs can be seen in Appendices 28a to 28v in Vol. 2 and the $t$-analysis of the Post-Study questionnaire data appears in Table 5.29, in Chapter 5.

**Student comments**

As mentioned previously, in addition to participants’ ratings of the assessment questionnaire items, which were analyzed quantitatively, the written comments of students were also of interest and could shed light on the ‘processes’ which had taken place in the students’ minds and their general attitudes to the whole assessment experience. For the Pre- and Main Studies, as explained previously, in Section 4.3.2, more precise qualitative data was drawn from the three open questions on the second page of the questionnaire (see second page of Appendix 21) and comments which had been written in English were recorded electronically and comments in Greek were translated into English by the I-R, recorded in Word documents and checked by an experienced translator.

In order to endeavour to give a more complete picture of responses to the assessment questionnaires, a few selected comments from students which were of relevance to each particular question have been inserted in the space after the analytical commentaries following each $xyz$ graph in Appendices 26a - j and 27a - j.

In the case of the Post-Study, space had been left after each of the 22 questions for comments, when students so wished. All of these were written in English and those which have been randomly selected to accompany the $xyz$ graphs, in Appendices 28a-v, are presented in unedited form. These comments give some substance to the statistics represented in the graphs.
4.8 Concluding remarks

The hypothesis and the four research questions on which the research methodology was based were outlined in this chapter and information was provided about the research design, the research instruments, the general background to the AARP, the pedagogical and assessment procedures and the analysis of assessment and questionnaire data. The next chapter will present the results which emerged from the research procedures.
Chapter 5
AARP Preliminary (Pre-), Main and Post-Study Results

5.0 Introduction

This chapter will set out, in four sections, the results of the Assessment for Autonomy Research Project (AARP). The first three sections will elaborate on the three parts of the AARP, which were: a Preliminary Study (Pre-Study), the Main Study and a Post-Study. The fourth section will present an overview of assessment rating, where Self-assessment is compared with Instructor-assessment, Peer-assessment is compared with Instructor-assessment and Peer-assessment is compared with Self-assessment patterns, throughout the five-year duration of the project.

The results which will be presented for each of the three Studies, which were different stages in the AARP, are those obtained from the triangulated Self-assessment (S-A), Peer-assessment (P-A) and Instructor assessment (I-A) of two home writing assignments and of one in-class oral presentation, produced by each Project participant in the context of a 1st year EFL course in the SOE, AUTh, entitled Language Mastery I (LM I), details of which were outlined in Chapter 4.

This strictly quantitative data, derived from assessment procedures, gives us some indication of the ‘products’ of the AARP, and will be balanced by some qualitative data gleaned from assessment questionnaires, which can offer more insights into the ‘processes’ and the workings of the minds that produced those ‘products’. It is hoped, with the data derived from both these sources, to be able to answer the four research questions posited (see Chapter 1, Section 1.4) and to be able to prove the general hypothesis on which the thesis is based; that is to say, that there are degrees of autonomy and that such degrees of autonomy can be promoted through learner-centred assessment.
5.1 The AARP Preliminary Study (Pre-Study)

The aim of the research in the Preliminary Study (Pre-Study) was firstly to seek answers to the four research questions outlined in Chapter 1, Section 1.4, and, depending on the answers to these questions, it would be decided if the activities, assignment tasks, assessment procedures and instruments used were worthy of adoption and replication in the years that followed, and, if deemed that they could be adopted and replicated, whether any modifications or adjustments to them would be required.

5.1.1 Breakdown of participants in the AARP Pre-Study, 2005-2006

A Preliminary Study of the Assessment for Autonomy Research Project (AARP) was conducted during the academic year 2005-2006, with the Instructor-Researcher’s two groups of Language Mastery I as the sample groups in the first semester of that year. As mentioned previously, the aim of the Pre-Study was to pilot the teaching, learning and assessment procedures and to evaluate if they yielded reasonably satisfactory results which would be worthy of repetition and duplication in the years to follow and, above all, if the pre-cognitive, cognitive and post-cognitive processes involved in the production and assessment of oral and written assignments in cycles of Peer-, Self- and Instructor-assessment would bring recognisable benefits to learners and instructor alike and help to promote some degree of autonomy, the idea being to use the processes of peer-assessment and self-assessment as possible stepping-stones to greater reflection, self-awareness, self-direction and self-determination, encouraging greater ownership of learning and the language being learned, and thus promote greater autonomy as well as valuable lifelong learning skills.

The research groups used in this Preliminary Study on Language Mastery I, in the autumn semester of 2005-2006 were the two groups of Language Mastery I,
which, to maintain the anonymity of the participants, will be referred to as Groups A and B, which were assigned to the Instructor-Researcher on a random basis from a total of ten Language Mastery I groups. The breakdown of the two groups according to age, ethnic origin and language ability is shown below in Table 5.1.

Individual participation in oral and writing tasks fluctuated, as did non-compulsory attendance on the course, but the maximum number of participants in any one task from Group A was 27, while the maximum number of participants from Group B was 23. In both groups, the female gender was predominant, as is typical of the School of English student population as a whole, with 24 females and 3 males in Group A and 21 females and 2 males in Group B. Likewise, the ethnic breakdown was predominantly Greek/Greek Cypriot with 24 Greeks/Greek Cypriots and 3 of other nationalities in Group A and 21 Greeks/Cypriots and 2 of other nationalities in Group B. 21 of the participants in Group A were aged 17-19, while the remaining 6 were aged 20 or over, while in Group B, 17 of the participants were aged 17-19 yrs, with the remaining 6 aged 20 or over.

The level of English of the participating students, based on completion of the Oxford Placement Test (UCLES, 2001) revealed that in Group A, 1 was C2 level, 8 were C1 level, 10 were B2 level, 3 were B1 level, and one was A2 level. The level of the 4 remaining participants, due to late joining of the course and non-completion of the OPT, was unknown, but since they were all repeating the course, their level was most likely B2 or below. In Group B, 2 were found to be C2 level, 6 were C1 level, 8 were B2 level, 5 were B1 level, and 2 were A2 level. It should be mentioned that the I-R also collected information about previous individual attainment levels in the language as this provided valuable insights into the degree of 2nd language attrition suffered, particularly in the last two years of attendance of senior high
<table>
<thead>
<tr>
<th>Academic Year</th>
<th>Group</th>
<th>Participants</th>
<th>Female</th>
<th>Male</th>
<th>17-19</th>
<th>20 and over</th>
<th>Greek/Cyp.</th>
<th>Other nat.</th>
<th>LEVEL AT ENTRY</th>
<th>LEVEL PREVIOUSLY ATTAINED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C2</td>
<td>C1</td>
<td>B2</td>
<td>B1</td>
</tr>
<tr>
<td>2005-2006</td>
<td>A</td>
<td>27</td>
<td>24</td>
<td>3</td>
<td>21</td>
<td>6</td>
<td>24</td>
<td>3</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2005-2006</td>
<td>B</td>
<td>23</td>
<td>21</td>
<td>2</td>
<td>17</td>
<td>6</td>
<td>21</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td>50</td>
<td>45</td>
<td>5</td>
<td>38</td>
<td>12</td>
<td>45</td>
<td>5</td>
<td>3</td>
<td>14</td>
</tr>
</tbody>
</table>

Levels shown are CEFR levels. U = Unknown
school, in the case of Greek students, and was also an indicator of the level of achievement in English which had been lost and could possibly be regained. This information is given in the final columns of Table 5.1 above, under ‘Previous Level Attained’, where U = unknown.

5.1.2 AARP Pre-Study 1st Home Writing Assignment assessment results
The procedures and methodology used for the production and assessment of the 1st Home Writing Assignment have been outlined in Chapter 4, in general terms in Section 4.5 and, more specifically, in Section 4.6.1. In order to ascertain if the first attempt at assessment of writing had been in any sense successful, the triangulated data gathered from assessment of the 1st Home Writing Assignment was subjected to statistical analysis. 25 participants from Group A and 22 from Group B took part in the full cycle of S-A, P-A and I-A. The means from S-A, P-A and I-A, for each group separately, were subjected to One-Way Analysis of Variance (ANOVA), where probability reference value \( p = 0.05 \).

The results, as shown in Table 5.2 below, were surprisingly closely aligned, with the I-R awarding the highest mean grade in both Groups A and B, while the learners revealed a fair amount of restraint in rating, as shown by Peer-assessors, as well as modesty, on the part of Self-assessors, typical of similar studies from the Far East (Cheng & Warren, 1999; Saito & Fujita, 2004; Matsuno 2007; 2009).
Table 5.2  AARP Pre-Study Mean Scores and ANOVA results for 1st Writing Assignment

<table>
<thead>
<tr>
<th>Group A</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A</td>
<td>7.42</td>
<td>0.81</td>
<td>0.16</td>
<td>25</td>
<td>F(2,72)=1.477</td>
<td>0.2351</td>
</tr>
<tr>
<td>P-A</td>
<td>7.46</td>
<td>1.27</td>
<td>0.25</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.86</td>
<td>0.89</td>
<td>0.18</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group B</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A</td>
<td>7.76</td>
<td>0.78</td>
<td>0.17</td>
<td>22</td>
<td>F(2,63)= 2.429</td>
<td>0.0953</td>
</tr>
<tr>
<td>P-A</td>
<td>7.38</td>
<td>0.89</td>
<td>0.19</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>8.00</td>
<td>1.11</td>
<td>0.24</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S-A = Self-assessment  P-A = Peer-assessment  I-A = Instructor assessment

A comparison of the means of the three different forms of assessment showed great similarity between S-A, at 7.42 and P-A, at 7.46, in Group A, while I-A was 0.4 higher, at 7.86. In Group B, on the other hand, S-A, at 7.76, was slightly higher (0.38) than P-A, at 7.38, but again both were lower than I-A, with a mean value of 8.00, revealing greater severity than the instructor. With regard to Standard Deviations (SD)s, in Group A, the highest levels of SD were displayed by peer-assessors, with a level of 1.27, followed by the I-R with 0.89, while in Group B the highest level of SD was displayed by the I-R, with a level of 1.11, while Peer-assessors came next with an SD level of 0.89. This indicates that Peer-assessors in Group A were rating using a wider range of grades than Group B Peer-assessors, and similar to the range employed by the I-R with Group B.

One-way Analysis of Variance (ANOVA) revealed a $p$ value of 0.2351 in the case of Group A, which is considered not significant, while in the case of Group B the $p$ value was 0.0953, which is also considered to be not significant. This reveals alignment in rating between S-A, P-A and I-A. Pearson product-moment analysis produced no significant correlation coefficients; nevertheless, the results overall were very encouraging indeed, particularly for a first attempt at writing assessment.
5.1.3 AARP Pre-Study 2nd Home Writing Assignment assessment results

In Chapter 4, Section 4.5.2, the general procedures, and in Section 4.6.2, the specific procedures relating to the production and assessment of the 2nd Home Writing Assignment were outlined. For the purposes of establishing if the second attempt at writing assessment had been as successful as the first attempt apparently was, the data gathered from the S-A, P-A and I-A cycle for the 2nd Home Writing Assignment was analysed statistically. In all, 24 participants from Group A and 19 from Group B took part in the full cycle of S-A, P-A and I-A. The means from S-A, P-A and I-A, for each group separately, were subjected to One-Way Analysis of Variance (ANOVA), where probability reference value \( p = 0.05 \). The scores from the assessment process of the 2nd Home Writing Assignment were subjected to statistical analysis, the results of which can be seen in Table 5.3 below:

<table>
<thead>
<tr>
<th>Group A</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A</td>
<td>7.53</td>
<td>0.86</td>
<td>0.176</td>
<td>24</td>
<td>F(2,69)=0.549</td>
<td>0.580</td>
</tr>
<tr>
<td>P-A</td>
<td>7.30</td>
<td>1.62</td>
<td>0.333</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.17</td>
<td>0.96</td>
<td>0.196</td>
<td>24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group B</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A</td>
<td>8.07</td>
<td>0.93</td>
<td>0.213</td>
<td>19</td>
<td>F(2,54)= 2.984</td>
<td>0.059</td>
</tr>
<tr>
<td>P-A</td>
<td>7.11</td>
<td>1.48</td>
<td>0.340</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.52</td>
<td>1.20</td>
<td>0.276</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Compared with the 1st Writing Assignment, the number of participants in Group A has fallen from 25 to 24 and, at the same time, has fallen from 22 to 19 in Group B. The pattern of means has also changed, with the highest mean scores of 7.53 and 8.07 being awarded by Self-assessors in Groups A and B respectively, while in Group A the mean of the Peer-assessors comes next in hierarchy at 7.30, followed by the Instructor-assessor at 7.17, while in Group B we see a different assessment pattern as the I-A comes next in rank order, with 7.52, followed by the mean of P-A, with 7.11.
With regard to Standard Deviations (SD), in both Groups A and B, the highest levels of SD were displayed in P-A, with levels of 1.62 and 1.48 respectively, followed by I-A in both cases, with 0.96 and 1.20 respectively, while in both Groups A and B, the lowest level of SD was displayed in S-A, with levels of 0.86 and 0.93 respectively. Overall, the grading behaviour of Self-assessors shows greater leniency than for the 1st Writing Assignment. The I-R displays much greater severity in assessment compared with the 1st Writing Assignment, as one might expect for a more demanding assignment, but is still not as severe as the Peer-assessors in Group B.

One-way Analysis of Variance (ANOVA) revealed a \( p \) value of 0.580 in the case of Group A, which is considered not significant, while in the case of Group B the \( p \) value was 0.059, which is also considered not significant. Pearson product-moment analysis, as in the case of the 1st Writing Assignment Assessment, produced no significant correlation coefficients; nevertheless, the overall results reveal assessment alignment, which appears to be satisfactory.

### 5.1.4 AARP Pre-Study Oral Assignment assessment results

General information regarding the approach and procedures taken in the conducting of LM I oral presentations and their assessment can be found in Chapter 4, Section 4.5.2, with more specific information concerning Oral assessment given in Section 4.6.3.

In order to ascertain if attempts at Self- and Peer-assessment of speaking skills were successful and comparable with instructor assessment, the triangulated data for S-A, P-A and I-A was subjected to statistical analysis. 27 participants from Group A and 23 from Group B took part in the full cycle of S-A, P-A and I-A. The means from
S-A, P-A and I-A, for each group separately, were subjected to One-Way Analysis of Variance (ANOVA), where probability reference value $p = 0.05$.

The results, as shown in Table 5.4 below, were surprisingly closely aligned, with P-A producing the highest and remarkably similar mean grade in both Groups A and B, of 7.86 and 7.88 respectively, while S-A followed close behind with 7.63 and 7.65 respectively, whereas I-A revealed slightly more restraint in rating, with mean grades of 7.21 and 7.48 for Groups A and B respectively.

### Table 5.4  AARP Pre-Study Mean Scores and ANOVA results for Oral Assignment

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A</td>
<td>7.62</td>
<td>1.16</td>
<td>0.22283</td>
<td>27</td>
<td>F(2,78)=1.925</td>
<td>0.1527</td>
</tr>
<tr>
<td>P-A</td>
<td>7.86</td>
<td>1.02</td>
<td>0.19639</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.21</td>
<td>1.50</td>
<td>0.28971</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Group B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A</td>
<td>7.65</td>
<td>1.16</td>
<td>0.24291</td>
<td>23</td>
<td>F(2,66)= 0.795</td>
<td>0.4559</td>
</tr>
<tr>
<td>P-A</td>
<td>7.88</td>
<td>0.76</td>
<td>0.15876</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.48</td>
<td>1.28</td>
<td>0.26784</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S-A = Self-assessment  \quad P-A = Peer-assessment  \quad I-A = Instructor assessment

With regard to Standard Deviations (SD), in both Group A and Group B, the highest levels of SD were displayed in I-A, with levels of 1.50 and 1.28 respectively, followed in both cases by S-A with 1.16 and 1.16 respectively, while the lowest SDs were revealed by P-A with 1.02 for Group A and 0.76 for Group B respectively. This indicates that the I-R was using a wider range of grades than Self-assessors and a considerably wider range than Peer-assessors.

One-way Analysis of Variance (ANOVA) revealed a $p$ value of 0.1527 in the case of Group A, which is considered not significant, while in the case of Group B the $p$ value was 0.4559, which is also considered not significant. Further statistical analysis of oral assessment data, using Pairwise Relationships revealed significant correlation coefficients of 0.75 and 0.51 between P-A and I-A in Groups A and B.
respectively, where $p<0.05$ (see Vol. 2, Appendix 17). These correlations provide further evidence of alignment in assessment between Peer-assessors and the I-R, particularly in the case of Group A.

Taken overall, in the Oral assessment, although P-A comes out as the most generous, and S-A, in both groups follows P-A quite closely, nevertheless the severity of I-A, which seems slightly greater with Group A than with Group B, does not deviate too greatly from S-A and P-A, thus producing what appears to be a valid and fair assessment process with a pleasing and satisfactory outcome, particularly given the previously-mentioned correlations between P-A and I-A in both groups.

5.1.5 AARP Pre-Study Questionnaire results

For the purposes of collecting qualitative feedback, a questionnaire was distributed to participants in both groups. For reasons of continuity of previous research and consistency of data, the questionnaire used was one devised by Antonopoulou in 2002 (see Appendix 21), which had been utilised for gathering information in a previous joint project conducted in 2003-2004 on Self-, Peer- and Instructor-assessment, reported in Antonopoulou et al. (2008) and in Joycey et al. (2010). Since most students were Greek-speakers, they were given the Greek version of the questionnaire, but could choose to write comments either in Greek or English. There were a total of 53 respondents from the I-R’s two Language Mastery I Groups, in the Preliminary Study of 2005-2006, with 28 students replying from Group A and 25 from Group B. Each question was answered on a scale of 1 to 5, where 1 = Completely disagree, 2 = Tend to disagree, 3 = Undecided, 4 = Tend to agree and 5 = Completely agree.

Since not all of the 22 questions in the original questionnaire (see Appendix 22, for the English version) are relevant to the present research, for the purposes of
reporting the outcomes of the AARP, the 22 questions were reduced to 10 (see Appendix 23, for the abridged version).

Given in Vol. 2, Appendices 26a to 26j are statistical representations, in the form of xyz graphs, of the responses to the 10 questions as answered by participants in the Pre-Study, with commentaries on the findings for each question given below each graph, together with selected relevant comments from students in each group. In addition, a $t$-analysis of the mean scores produced by each group for each question (see Appendix 23) is given in Table 5.5 below, where $p = 0.05$:

<table>
<thead>
<tr>
<th>Question</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>28</td>
<td>3.61</td>
<td>0.786</td>
<td>t(51)=0.634</td>
<td>0.529</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>3.76</td>
<td>0.970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>28</td>
<td>3.64</td>
<td>0.870</td>
<td>t(51)=1.385</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>3.96</td>
<td>0.790</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>28</td>
<td>3.11</td>
<td>1.166</td>
<td>t(51)=2.639</td>
<td><strong>0.011</strong></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>3.84</td>
<td>0.800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>28</td>
<td>3.50</td>
<td>1.106</td>
<td>t(51)=2.293</td>
<td><strong>0.026</strong></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>4.08</td>
<td>0.640</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>28</td>
<td>2.57</td>
<td>1.168</td>
<td>t(51)=0.899</td>
<td>0.373</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>2.84</td>
<td>0.987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>28</td>
<td>3.29</td>
<td>0.976</td>
<td>t(51)=1.074</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>3.56</td>
<td>0.870</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>28</td>
<td>3.36</td>
<td>1.096</td>
<td>t(51)=2.277</td>
<td><strong>0.027</strong></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>3.96</td>
<td>0.790</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>28</td>
<td>3.61</td>
<td>1.133</td>
<td>t(51)=0.854</td>
<td>0.397</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>3.84</td>
<td>0.800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>28</td>
<td>3.86</td>
<td>1.044</td>
<td>t(51)=1.433</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>4.24</td>
<td>0.879</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>28</td>
<td>3.39</td>
<td>0.916</td>
<td>t(51)=1.879</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>25</td>
<td>3.84</td>
<td>0.800</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What is interesting from the $t$-analysis of the mean scores of the responses to the questionnaires from Groups A and B is the fact that for all questions, a higher mean score is awarded by Group B, showing what seems to be a generally more positive
attitude, on their part, to assessment procedures and the experience and understanding derived from those procedures. This positive attitude results in significant differences of opinion on three out of the ten questions, for which the $p$ values are highlighted in bold in Table 5.5, namely Questions 3, 4 and 7, which were related to how easy it was to peer-assess and how easy it was to be objective in peer-assessment, while Question 7 was concerned with the usefulness of the criteria check-lists in understanding how learners were being assessed. Clearly, Group B felt more comfortable with peer-assessment and with the checklists than Group A, even if this does contradict to some extent the correlation coefficients produced between P-A and I-A in the Oral assessment. A slightly more positive attitude towards Peer-assessment, and a willingness to try, also seems to come across in written comments from members of Group B (see Appendices 26c & 26d), while the comments about the usefulness of checklists reveal neutrality on the part of both Group A and B members (see Appendix 26g).

5.1.6 AARP Pre-Study - Overview of outcome

The four research questions on which this thesis is based are outlined in Chapter 1, Section 1.4. Judging from the results of the AARP Pre-Study, outlined in this chapter, Sections 5.1.2, 5.1.3 and 5.1.4, which were related to self-assessment and peer-assessment of 2 writing assignments and 1 oral assignment on the course, LM I, the first three of our four research questions seem to have been answered. There appears to be clear evidence that in all assignments, the participants used the criteria check-lists provided, and used them, both in the case of writing skills and in the case of speaking skills, with objectivity and reliability. This is significant because the participants received no formal training in assessment, had no previous experience of peer- and self-assessment, and were able to self-assess based on their experience of
peer-assessment, which therefore proved to be a satisfactory stepping-stone, on the path to self-assessment, and on the road to what Falchikov (2007, p. 139) refers to as “assessment autonomy”. This also indicates that the criteria check-lists, although not negotiated with the participants, were adopted by them and used in a manner similar to the I-R, thus following the same frame of reference and displaying both cooperation and trust.

Although there are some instances in the questionnaire data, presented in Section 5.1.5, where one group is more hesitant than the other in its response to a question and other instances where one group takes a more positive or more negative stance (see also Appendices 26a to 26j), overall there is a similarity in the two groups’ impressions of and attitude towards peer- and self-assessment and their value, and towards the usefulness of the assessment criteria check-lists, giving the impression that the AARP has been a valuable learning experience for both Groups A and B. Thus, the reply to the 4th research question (see Chapter 1, Section 1.4), with regard to the Pre-Study, appears to be affirmative.

It can also be said that all indications, both quantitative and qualitative, seemed to show that the 2005-2006 AARP Pre-Study would be worthy of replication, which is what occurred in the 3-4 years which followed.

5.2 The AARP Main Study Results
The aim in this section will be to record the results of the AARP Main Study, which took place in the autumn semesters of the academic years, 2006-2009 in the SOE, AUTh. The Main Study was conducted with students from 1st Year Language Mastery I courses and, as in the case of the AARP Pre-Study, involved Self-, Peer- and Instructor assessment of 2 home writing assignments and 1 oral presentation,
given before a whole-class audience. Information related to course procedures, tasks and assessment processes have already been outlined, in detail, in Chapter 4.

Having established in the AARP Pre-Study in the academic year 2005-2006 that all procedures, tasks and research instruments were satisfactory, the same methods and methodology were replicated in the three years of the Main Study that followed. The aims of the AARP Main Study were essentially to find answers to the four research questions on which the hypothesis of the thesis is based, which were outlined in Chapter 1, Section 1.4. Results obtained from the AARP Main Study will be set out and subsequently be compared with those of the AARP Pre-Study.

5.2.1 Breakdown of participants in the AARP Main Study, 2006-2009
As previously mentioned, the AARP Main Study was conducted in the autumn semester of the three academic years between 2006 and 2009. Again, it was the groups assigned to the I-R to teach in her place of employment, in the SOE, AUTh, which were used as the Research Groups in each case and were therefore convenience samples.

As in the Pre-Study, participants were given the Oxford Placement Test (UCLES, 2001) to ascertain their level of English on entry to their LM I courses and to the AARP. Details of age and ethnic origin were also recorded so that any changes in variables which might affect the outcome of the research could be noted. Some details about the participants are shown in Table 5.6 below.

As was the case with the Pre-Study, in the Main Study the number of participants from any one particular group fluctuated from task to task, but the maximum number of participants in Groups C and D in 2006-2007 were 23 and 18 respectively, in Groups E and F in 2007-2008, were 30 and 23 respectively, and in Groups G and H in 2008-2009, were, likewise, 30 and 23 respectively. In all six
groups, the female gender predominated as is the case with the departmental student population as a whole, with 20 females and 3 males and 16 females and 2 males in Groups C and D, 2006-2007, respectively, 27 females and 3 males and 19 females and 4 males in Groups E and F, 2007-2008, respectively, 25 females and 5 males and 21 females and 2 males in Groups G and H, 2008-2009, respectively.

In terms of nationality, Greeks and Greek Cypriots predominated, with 23 Greeks and 18 Greeks and no foreigners in Groups C and D, 2006-2007 respectively, with 28 Greeks and 2 foreigners and 23 Greeks and no foreigners in Groups E and F, 2007-2008 respectively, and with 30 Greeks and no foreigners and 21 Greeks and 2 foreigners in Groups G and H, 2008-2009, respectively.

With regard to age, 20 were aged 17-19 yrs of age and 3 were aged 20 or over and 15 were aged 17-19 yrs and 3 were aged 20 or over in Groups C and D, 2006-2007, respectively, while 20 were aged 17-19 yrs and 10 were aged 20 or over, and 17 were aged 17-19 yrs and 6 were aged 20 yrs or over in Groups E and F, 2007-2008, respectively, and, finally, 18 were aged 17-19 yrs and 12 were aged 20 or over and 17 were aged 17-19 yrs and 6 were aged 20 or over in Groups G and H, 2009-2010, respectively.

Concerning language ability, based on completion of the OPT (UCLES, 2001), the participants in the six research groups in the Main Study were found to lack uniformity regarding their CEFR levels of knowledge of the language. Group C in 2006-2007 was found to have no students at C2 level, but had 6 at C1 level, 15 at B2 level, 1 at B1 level and no students at A2 level, while Group D in the same year also
Table 5.6   AARP Main Study, 2006-2009 - Group breakdown

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>Group</th>
<th>Participants</th>
<th>Female</th>
<th>Male</th>
<th>17-19</th>
<th>20 and over</th>
<th>Greek/Cyp.</th>
<th>Other nat.</th>
<th>LEVEL AT ENTRY</th>
<th>LEVEL PREVIOUSLY ATTAINED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C2</td>
<td>C1</td>
</tr>
<tr>
<td>2006-2007</td>
<td>C</td>
<td>23</td>
<td>20</td>
<td>3</td>
<td>20</td>
<td>3</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>2006-2007</td>
<td>D</td>
<td>18</td>
<td>16</td>
<td>2</td>
<td>15</td>
<td>3</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2007-2008</td>
<td>E</td>
<td>30</td>
<td>27</td>
<td>3</td>
<td>20</td>
<td>10</td>
<td>28</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>2007-2008</td>
<td>F</td>
<td>23</td>
<td>19</td>
<td>4</td>
<td>17</td>
<td>6</td>
<td>23</td>
<td>0</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>2008-2009</td>
<td>G</td>
<td>30</td>
<td>25</td>
<td>5</td>
<td>18</td>
<td>12</td>
<td>30</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>2008-2009</td>
<td>H</td>
<td>23</td>
<td>21</td>
<td>2</td>
<td>17</td>
<td>6</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td>147</td>
<td>128</td>
<td>19</td>
<td>107</td>
<td>40</td>
<td>143</td>
<td>4</td>
<td>9</td>
<td>43</td>
</tr>
</tbody>
</table>

Levels shown are CEFR levels. U = Unknown
had no students at C2 level, 3 at C1 level, 8 at B2 level, 6 at B1 level and none at A2 level. From the Research Groups in 2007-2008, Group E was found to have 2 students at C2 level, 9 students at C1 level, 14 students at B2 level, 3 students at B1 level, and no students at A2 level, while in Group F in the same semester, 4 students were found to be at C2 level, 8 students at C1 level, 6 students at B2 level, 1 student at B1 level and 2 students at A2 level. Of the two participating groups in 2008-2009, Group G was found to have 2 students at C2 level, 10 students at C1 level, 11 students at B2 level, 2 students at B1 level and no students at A2 level, while Group H had 1 student at C2 level, 7 students at C1 level, 9 students at B2 level, 1 student at B1 level and no students at A2 level. Information was also gathered about previous attainment levels in the language in order to understand the degree of attrition that had occurred with particular individuals.

5.2.2 AARP Main Study 1st Home Writing Assignment assessment results
The same procedures, which are described in Chapter 4, Sections 4.5 and 4.6.1, were followed in the Main Study, as in the Pre-Study, regarding the 1st Home Writing Assignment and the data was collected and was collated in the exact same way. The means derived from S-A, P-A, and I-A, in each group, were subjected to One-Way Analysis of Variance (ANOVA), where probability reference value $p = 0.05$. In cases where significant differences were found, further analysis using the Tukey-Kramer Comparison Test of Pairwise Mean Differences was deployed. The data was processed and analysed statistically and produced the following, as shown below in Table 5.7:
Table 5.7 AARP Main Study Mean Scores and ANOVA results for 1st Writing Assignment

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007 Group C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A</td>
<td>8.67</td>
<td>0.85</td>
<td>0.172</td>
<td>24</td>
<td>F(2,69)=6.809</td>
<td>0.002</td>
</tr>
<tr>
<td>P-A</td>
<td>7.28</td>
<td>1.81</td>
<td>0.369</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.72</td>
<td>1.22</td>
<td>0.249</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007 Group D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A</td>
<td>7.61</td>
<td>1.26</td>
<td>0.316</td>
<td>16</td>
<td>F(2,45)=0.541</td>
<td>0.586</td>
</tr>
<tr>
<td>P-A</td>
<td>7.16</td>
<td>1.32</td>
<td>0.331</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.46</td>
<td>1.20</td>
<td>0.300</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2008 Group E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A</td>
<td>8.49</td>
<td>0.77</td>
<td>0.145</td>
<td>28</td>
<td>F(2,81)=5.453</td>
<td>0.006</td>
</tr>
<tr>
<td>P-A</td>
<td>7.96</td>
<td>1.40</td>
<td>0.263</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.50</td>
<td>1.11</td>
<td>0.209</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2008 Group F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A</td>
<td>8.50</td>
<td>0.96</td>
<td>0.214</td>
<td>20</td>
<td>F(2,57)=1.785</td>
<td>0.177</td>
</tr>
<tr>
<td>P-A</td>
<td>7.87</td>
<td>1.19</td>
<td>0.266</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.97</td>
<td>1.23</td>
<td>0.275</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009 Group G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A</td>
<td>8.27</td>
<td>0.76</td>
<td>0.147</td>
<td>27</td>
<td>F(2,78)=5.675</td>
<td>0.005</td>
</tr>
<tr>
<td>P-A</td>
<td>7.36</td>
<td>1.24</td>
<td>0.238</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.76</td>
<td>0.93</td>
<td>0.179</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009 Group H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A</td>
<td>8.37</td>
<td>1.13</td>
<td>0.241</td>
<td>22</td>
<td>F(2,63)=2.103</td>
<td>0.130</td>
</tr>
<tr>
<td>P-A</td>
<td>7.75</td>
<td>0.99</td>
<td>0.210</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.89</td>
<td>1.06</td>
<td>0.226</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S-A = Self-assessment    P-A = Peer-assessment    I-A = Instructor assessment

If we compare the means produced by S-A, P-A and I-A in the Main Study with those of the Preliminary Study, the first thing worthy of note is the shifting in all cases from I-A having the highest means to having the next-highest means, in all but one [Group E, 2007-2008] of the six research groups. In all but two cases [Group F, 2007-2008 and Group H, 2008-2009], the highest SD applies to P-A, and out of those four, the highest value occurs in Group C, 2006-2007, with an SD value of 1.81. This indicates
that Peer-assessors were awarding a wider range of grades than either Self-Assessors or the I-R.

In Group C (2006-2007), Group E (2007-2008) and Group G (2008-2009), ANOVA revealed \( p \) values of 0.002, 0.006 and 0.005 respectively which are considered to be significant (see Table 5.8 below), while in Group D in 2006-2007, Group F in 2007-2008 and Group G in 2008-2009, One-Way ANOVA revealed \( p \) values of 0.586, 0.177 and 0.130 respectively, which are considered not to be significant (see Table 5.8 below). Tukey-Kramer Multiple Comparison Tests conducted on Group C in 2006-2007, Group E in 2007-2008 and Group G in 2008-2009 produced the pattern of A>C=B, A>C only and A>B only, respectively, showing that there was no consistency (see Table 5.8 below).

Table 5.8  AARP Main Study 1st Writing assignment assessment – Tukey-Kramer mean comparisons and Pearson correlation coefficients

<table>
<thead>
<tr>
<th>Year</th>
<th>Group</th>
<th>( n )</th>
<th>One-way ANOVA &amp; Tukey-Kramer</th>
<th>P-I Pearson correlation coefficient</th>
<th>S-I Pearson correlation coefficient</th>
<th>P-S Pearson correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007</td>
<td>Group C</td>
<td>24</td>
<td>sig. ( p = 0.002 ) A&gt;C=B</td>
<td>0.62</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>2006-2007</td>
<td>Group D</td>
<td>16</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td>Group E</td>
<td>28</td>
<td>sig. ( p = 0.006 ) A&gt;C only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td>Group F</td>
<td>20</td>
<td>n.s.</td>
<td>0.51</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td>Group G</td>
<td>27</td>
<td>sig. ( p = 0.005 ) A&gt;B only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td>Group H</td>
<td>22</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n.s. = non-significant     sig. = significant   A = Self    S-I = Self-Instructor
B = Peer                  P-I = Peer-Instructor
C = Instructor-researcher  P-S = Peer-Self

Analysis of Pairwise Relationships also revealed statistically significant correlation coefficients between S-A and I-A, with a value of 0.61, and between P-A and I-A, with a value of 0.62 in Group C in the year 2006-2007, while a correlation coefficient of 0.50 between S-A and P-A and of 0.51 between S-A and I-A also occurred in Group F, in the year 2007-2008. In the year 2008-2009, there were no
statistically significant correlations in either Group G or Group H. Taken overall, the results for Groups E and G come across as the most disappointing.

5.2.3 AARP Main Study 2nd Home Writing Assignment assessment results
The same pedagogical procedures and assessment processes were followed in the Main Study as in the Pre-Study for the 2nd Home Writing Assignment, described in Chapter 4, in general terms, in Section 4.5, and specific terms, in Section 4.6.2, with the exception of Groups C and D, where a variation in Self-assessment procedures, and its significance, will be described and reported in Section 5.2.3.1 below. With the four remaining groups, the same procedures concerning preparation, submission and triangulated assessment were followed as in the Pre-Study.

As in the case of the 1st Home Writing Assignment, the Means for S-A, P-A and I-A for each group were calculated and subjected to One-Way Analysis of Variance (ANOVA), where probability reference value $p = 0.05$. In cases where significant differences were found, further analysis using the Tukey-Kramer Comparison Test of Pairwise Mean Differences was deployed. The collated data was subjected to statistical analysis, which produced the following, as shown in Table 5.9:

<table>
<thead>
<tr>
<th>2006-2007 Group C</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A</td>
<td>8.80</td>
<td>0.67</td>
<td>0.142</td>
<td>22</td>
<td>F(2,63)=8.365</td>
<td>0.0006</td>
</tr>
<tr>
<td>P-A</td>
<td>7.94</td>
<td>1.12</td>
<td>0.240</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.42</td>
<td>0.81</td>
<td>0.174</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2006-2007 Group D</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A</td>
<td>8.44</td>
<td>1.05</td>
<td>0.248</td>
<td>18</td>
<td>F(2,51)=5.047</td>
<td>0.010</td>
</tr>
<tr>
<td>P-A</td>
<td>7.41</td>
<td>1.26</td>
<td>0.297</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.41</td>
<td>1.03</td>
<td>0.242</td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A comparison of the means of the three different forms of assessment showed similarity between the two research groups in each of the years 2006-2007, 2007-2008 and 2008-2009, in the Main Study, in that the highest mean value in each case was always produced by S-A, with values of 8.80, 8.44, 8.77, 8.73, 8.59 and 8.60 respectively, all of which are well beyond the S-A values of 7.53 and 8.07 produced by the two research groups in the Preliminary Study of 2005-2006. In 4 out of 6 cases, the P-A value was the next highest mean, while in one case, in Group D in 2006-2007, it was equal with I-A, with a value of 7.41 and in the other, in Group F, 2008-2009, the I-A was significantly higher than P-A, with values of 8.16 and 7.61 respectively, indicating, in this instance, greater leniency on the part of the I-R.

With regard to Standard Deviation (SD), the highest level of SD was displayed by P-A in 4 out of 6 cases, with levels of 1.12, 1.26, 1.04 and 1.14 respectively, indicating a wider range of grades being awarded by Peer-assessors, while with the exception of Group F, 2007-2008, the I-A SD value was the highest, with 1.34,
followed by P-A, with 1.12 and S-A, with 0.88, while in the other exception, in Group H, 2008-2009, the I-A SD value was again the highest, with 1.25, but this time was followed by S-A, with 1.12 and P-A, with 0.98, showing greater similarity between Self-assessors and the I-R in the range of grades being awarded in this instance.

One-way ANOVA revealed significant $p$ values in all research groups (see Table 5.9 above), with the exception of Group F, 2007-2008, where the $p$ value of 0.104 is considered not significant. Tukey-Kramer Multiple Comparison Tests (see Table 5.10 below) revealed very similar results in Groups C and D, 2006-2007 and Group E in 2007-2008, with the pattern A>B=C, an indication of greater alignment between Peer-assessment and Instructor-assessment. In Group G, 2008-2009, because of the marginal $p$ value, no differences can be detected, while in Group H of the same year, A>B only.

Analysis of Pairwise Relationships revealed statistically significant correlation coefficients in Group C, 2006-2007 between S-A and I-A, with a value of 0.48, while a very similar result occurred in Group G, 2008-2009, between S-A and I-A, with a value of 0.47, while the same occurred in Group H, 2008-2009, between S-A and I-A, but with a slightly higher value of 0.66, all of which seem to be indications of some degree of agreement between S-A and I-A. No such correlations were found to occur in the year 2007-2008, as shown in Table 5.10 below:
The results derived from analysis of the assessment data for the 2nd Writing assignment in the Main Study are altogether different from those of the Pre-Study, but nevertheless contain some interesting indications of alignment, both between P-A and I-A and between S-A and I-A.

### 5.2.3.1 AARP Main Study (2006-2007) Variation in Self-Assessment of the 2nd Writing Assignment

Previously, in Section 5.2, it was mentioned that in Groups C and D, 2006-2007, in the Main Study, that there was a unique variation in the way Self-assessment of the 2nd Writing Assignment was conducted and details of this variation in procedures were outlined in Chapter 4, Section 4.6.2.1, concerning the way that self-assessment of the assignment was conducted, which was twice-over: (1) on submission of the assignment and (2) in the usual way, after peer-assessment had been conducted.

This variation in procedures has serious implications for the assessment results presented in Table 5.9, for the academic year 2006-2007, because, as will be revealed, things are not exactly as they seem. The reason for this is that not all of the students from Groups C and D participated in this variation, so that only 20 of the 22 participants in assessment processes in Group C and, still fewer, with only 13 of the...
18 participants from Group D, were involved; nevertheless, it seems important to report the outcome of this particular experiment. Table 5.11 below shows the results from statistical analysis of these variations, where:

(1) ‘Actual’ refers to the means derived from all the participants who were involved in Self-assessment after conducting Peer-assessment. These means are referred to as ‘Actual’ because they are the same as those presented in Table 5.9 above;

(2) ‘Variation 1’ presents the means derived from learner-assessment for those students, fewer in number, from Groups C and D, who submitted Self-assessment with their assignments, before Peer-assessment processes took place. The S-A means presented in Variation 1 are therefore those based on their first Self-assessment, before involvement in Peer-assessment;

(3) ‘Variation 2’ presents the same means as in ‘Actual’, but with means derived only from those students who performed Self-assessment twice, i.e., only those students who took part in Variation 1.

These three sets of means are presented in Table 5.11, below:

<table>
<thead>
<tr>
<th>Table 5.11 AARP Main Study Mean Scores and ANOVA results for 2nd Writing Assignment S-A Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007 Group C (Actual)</td>
</tr>
<tr>
<td>S-A</td>
</tr>
<tr>
<td>P-A</td>
</tr>
<tr>
<td>I-A</td>
</tr>
<tr>
<td>2006-2007 Group C (Variation 1)</td>
</tr>
<tr>
<td>S-A</td>
</tr>
<tr>
<td>P-A</td>
</tr>
<tr>
<td>I-A</td>
</tr>
<tr>
<td>2006-2007 Group C (Variation 2)</td>
</tr>
<tr>
<td>S-A</td>
</tr>
<tr>
<td>P-A</td>
</tr>
<tr>
<td>I-A</td>
</tr>
<tr>
<td>2006-2007 Group D (Actual)</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>S-A</td>
</tr>
<tr>
<td>P-A</td>
</tr>
<tr>
<td>I-A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2006-2007 Group D (Variation 1)</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A</td>
<td>8.11</td>
<td>1.19</td>
<td>0.329</td>
<td>13</td>
<td>F(2,36)=2.023</td>
<td>0.147</td>
</tr>
<tr>
<td>P-A</td>
<td>7.29</td>
<td>1.31</td>
<td>0.362</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.29</td>
<td>1.08</td>
<td>0.300</td>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2006-2007 Group D (Variation 2)</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A</td>
<td>8.71</td>
<td>1.08</td>
<td>0.300</td>
<td>13</td>
<td>F(2,36)=6.462</td>
<td>0.004</td>
</tr>
<tr>
<td>P-A</td>
<td>7.29</td>
<td>1.31</td>
<td>0.362</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A</td>
<td>7.29</td>
<td>1.08</td>
<td>0.300</td>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As previously mentioned, Table 5.11 above gives a comparison of the means of the three different forms of assessment S-A, P-A and I-A for Groups C and D of the Main Study, 2006-2007. Parts of the table are subtitled ‘Actual’, ‘Variation 1’ and ‘Variation 2’, as appropriate, to allow for comparisons to be made. The highest means in all cases, whether Actual, Variation 1 or Variation 2 for both Groups C and D was produced by S-A, with higher S-A values in each case awarded by Group C. With the exception of Group D (Variation 1), all of the S-A values are well beyond the values of 7.53 and 8.07 produced by the two research groups in the Preliminary Study of 2005-2006 (see Table 5.3 above). Group D (Variation 1) has an S-A value of 8.11 which comes close to the S-A value of Group B, as just mentioned, in the Preliminary Study. In Group C, the P-A value was the next highest mean, followed by I-A, both in the Actual and in Variations 1 and 2, while in the case of Group D, the equivalent P-A and I-A values are equal, both in the Actual and in Variations 1 and 2, revealing consistent closeness in Peer- and Instructor-assessment.
With regard to Standard Deviation (SD), the highest level of SD was displayed by P-A in all cases, indicating that Peers were awarding a wider range of grades than both Self-assessors and the I-R. One-way ANOVA revealed significant $p$ values for both Groups C and D in the Actual and in Variation 2, but in the case of Variation 1, while Group C still produced a $p$ value of 0.002, which is considered significant, in the case of Group D, the $p$ value was 0.147, which is considered non-significant. Tukey-Kramer Multiple Comparison Tests revealed very similar results in Groups C and D (Actual) and Groups C and D (Variation 2), with the pattern A>B=C, indicating alignment between Peer-assessment and Instructor-assessment.

In order to understand better the differences in assessment behaviour between Variation 1 and Variation 2, a paired sample t-test of the two self-assessments of the two groups was conducted and revealed the following, as shown in Table 5.12:

| Table 5.12  Paired t-test Results for 2nd Writing Assignment Self-assessment Variations (2006-2007) |
|---------------------------------|---|---|---|---|
|                                | n | Mean Difference | SD  | SEM |
| **2006-2007**                  |   |                |     |     |
| Group C                        | 20 | 0.33           | 0.54| 0.121|
| 95% CI for mean difference:  (0.077; 0.583) |
| T-Test of mean difference = 0 (vs not=0); T-Value = 2.73  P-value= 0.013 |
| n | Mean Difference | SD  | SEM |
| **2006-2007**                  |   |                |     |     |
| Group D                        | 13 | 0.60           | 0.63| 0.17 |
| 95% CI for mean difference:  (0.221; 0.979) |
| T-Test of mean difference = 0 (vs not=0); T-Value = 3.45  P-value= 0.005 |

Both the ANOVA and the t-test reveal significant increases in S-A values for both Groups C and D, with the Paired t-test revealing an increase of 0.33 for Group C and an increase of 0.60 for Group D between the first Self-assessment conducted and the second. These increases occur after peer-assessment processes, which we would have expected to have been a form of training in using the criteria and to have had more of a regulatory effect on consequent self-assessment processes, while, on the contrary,
peer-assessment appears to have had a negative effect on self-assessment, since S-A means have increased the second time around.

Analysis of Pairwise Relationships revealed statistically significant correlation coefficients (see Table 5.13 below) in Group C, 2006-2007 (Actual) between S-A and I-A, with a value of $r = 0.48$ and in Group D, 2006-2007 (Variation 1) between S-A and I-A, with a value of $r = 0.62$, indicating a tendency towards alignment in these cases.

Table 5.13  AARP Main Study 2nd Writing Assignment assessment Variations – Tukey-Kramer mean comparisons and Pearson correlation coefficients

<table>
<thead>
<tr>
<th>$p=0.05$</th>
<th>n</th>
<th>One-way ANOVA &amp; Tukey-Kramer</th>
<th>P-I Pearson correlation coefficient</th>
<th>S-I Pearson correlation coefficient</th>
<th>P-S Pearson correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007 Group C (ACTUAL)</td>
<td>22</td>
<td>sig. $p &lt; 0.001$ A&gt;B=C</td>
<td></td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td>2006-2007 Group C (VARIATION 1)</td>
<td>20</td>
<td>sig. $p = 0.002$ A&gt;B only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007 Group C (VARIATION 2)</td>
<td>20</td>
<td>sig. $p &lt; 0.001$ A&gt;B=C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007 Group D (ACTUAL)</td>
<td>18</td>
<td>sig. $p = 0.010$ A&gt;B=C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007 Group D (VARIATION 1)</td>
<td>13</td>
<td>n.s. $p = 0.147$</td>
<td></td>
<td></td>
<td>0.62</td>
</tr>
<tr>
<td>2006-2007 Group D (VARIATION 2)</td>
<td>13</td>
<td>sig. $p = 0.004$ A&gt;B=C</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n.s. = non-significant  sig. = significant  A = Self  B = Peer  C = Instructor-researcher  S-I = Self-Instructor  P-I = Peer-Instructor  P-S = Peer-Self

Looking at Table 5.13 above, it is, of course, difficult to reach any conclusions based on one case of no significant differences between the means, as in Group D, Variation 1, and on two correlation coefficients, one of these produced in the same aforementioned instance and the other in Group C, Actual. Comparisons are particularly difficult when Group D, with only 13 participants in Variation 1, was so
much smaller than Group C, with 22 participants, in what is being referred to as Actual.

It is, however, extremely interesting that Group D in Variation 1, i.e., before peer-assessment processes, produced a greater S-I correlation coefficient, of 0.62, than that for Group C (Actual), which was 0.48, as seen in Table 5.13 above, and it is also very interesting that the Means for P-A and I-A are equal in Group D, both in Actual and in the Variations (see Table 5.11 above). This could be seen as an indication, together with the t-test results provided in Table 5.12, though this is expressed with great caution, that in Group D (though the sample is small), and possibly also in Group C, there was sufficient assessment “maturity” (Ritter, 1998, p. 79), before peer-assessment processes, for these students to self-assess themselves reasonably accurately, indicating that practice through peer-assessment with these groups perhaps led to less rather than more self-assessment accuracy. It is otherwise hard to explain why students had the tendency to overrate in self-assessment processes, in Groups C and D, to the extent that they did, after an experience of peer-assessment which had proved to be rather successful.

More will be mentioned about rating behaviour on the 2nd Writing Assignment in Section 5.4.2, and the deviant behaviour of Groups C and D, combined, is illustrated in Figure 5.2 in the same chapter section.

5.2.4 AARP Main Study Oral Assignment assessment results
The educational procedures and the assessment processes which were followed in the case of Oral Skills on Language Mastery I courses have already been described in Chapter 4, in Section 4.5.2, in general terms, and in Section 4.6.3, in specific terms, and the overall significance of the Oral Presentations in terms of weighting in the final course grade are discussed there.
For assessment purposes the same form of triangulated assessment between P-A, S-A and I-A was selected and the same set of criteria (see Appendix 10) which had seemed to work effectively in the Pre-Study, were utilised in the Main Study. Details of what the presentations entailed and the instrumentation used in Self-, Peer- and Instructor-assessment of the presentations are outlined in Chapter 4, Section 4.6.3.

In the case of Oral assessment, each member of the audience participated in the Peer-assessment process and the presenters themselves conducted Self-assessment on completion of their presentations, as did the Instructor. At a later date, the multiple peer grades were averaged out and the grades for S-A, P-A and I-A were recorded, both on Profile Cards (see Appendix 3) and electronically. After recording, the Means for S-A, P-A and I-A were calculated for each group and were subjected to One-Way Analysis of Variance (ANOVA), where the probability reference value $p = 0.05$. In those cases where significant differences were found, further analysis using the Tukey-Kramer Comparison Test of Pairwise Mean Differences was deployed. The Mean Score values produced by statistical analysis for groups in the Main Study are shown in Table 5.14 below:

<table>
<thead>
<tr>
<th>Table 5.14</th>
<th>AARP Main Study Mean Scores and ANOVA results for Oral Assignment assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007 Group C</td>
<td>Mean</td>
</tr>
<tr>
<td>S-A</td>
<td>8.82</td>
</tr>
<tr>
<td>P-A</td>
<td>8.37</td>
</tr>
<tr>
<td>I-A</td>
<td>8.23</td>
</tr>
<tr>
<td>2006-2007 Group D</td>
<td>Mean</td>
</tr>
<tr>
<td>S-A</td>
<td>8.27</td>
</tr>
<tr>
<td>P-A</td>
<td>8.56</td>
</tr>
<tr>
<td>I-A</td>
<td>7.73</td>
</tr>
<tr>
<td>2007-2008 Group E</td>
<td>Mean</td>
</tr>
<tr>
<td>S-A</td>
<td>8.75</td>
</tr>
<tr>
<td>P-A</td>
<td>8.39</td>
</tr>
<tr>
<td>I-A</td>
<td>7.87</td>
</tr>
</tbody>
</table>

193
A comparison of the means of the three different forms of assessment reveals that in each of the six research groups involved in the three-year study, with only one exception, S-A comes out with the highest mean value in a range between 8.40 and 8.80. In the exceptional case, which is Group D, 2006-2007, there is a -0.29 Mean Difference between S-A and P-A, with S-A having a value of 8.27 and P-A, a value of 8.56 respectively. In the remaining 5 cases, the P-A mean is the next highest value, ranging between 8.00 and 8.30, with the exception of Group H, 2008-2009, where the P-A value is 8.23 and falls behind the value of I-A, with 8.53. Then, as we would expect from what has just been stated, I-A, with the exception of Group H 2008-2009, is always the lowest mean value, with a range between 7.70 and 8.20.

One-way ANOVA (see Table 5.14 above) revealed $p$ values of 0.086 and 0.068, both considered not significant, for Groups C and D, 2006-2007, and $p$ values of 0.0007 and 0.119, considered significant in the former case and not significant in the latter, for Groups E and F, 2007-2008, and $p$ values of 0.062 and 0.384, both considered not significant, for Groups G and H, 2008-2009, respectively. In the exceptional case of Group E, 2007-2008, a Tukey-Kramer Multiple Comparison Test...
revealed the pattern $A=B>C$, which breaks with the norm, as can be seen in Table 5.15 below.

Analysis of Paired Relationships revealed significant correlation coefficients (see Table 5.15 below) between $P-A$ and $S-A$ ($P-S$), with similar values of 0.67 and 0.65, respectively, in Groups C and D, 2006-2007, showing fairly strong alignment in assessment behaviour between Peer- and Self-assessors. Likewise, in Group C only, there was a correlation between $S-A$ and $I-A$ ($S-I$), with a value of 0.45, which is encouraging. In the following year, 2007-2008, there is a similar pattern of correlation coefficients between $P-A$ and $S-A$ ($P-S$), with values of 0.50 and 0.44 in Groups E and F respectively. Likewise, in Group F this time, there is a correlation between $S-A$ and $I-A$ ($S-I$), with a value of 0.42, as well as a correlation between $P-A$ and $I-A$ ($P-I$), with a coefficient value of 0.46. While these correlations may not be particularly strong, they do show particular trends or tendencies towards assessment alignment, which is pleasing.

Table 5.15 AARP Main Study Oral assignment assessment – Tukey-Kramer mean comparisons and Pearson correlation coefficients

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>One-way ANOVA &amp; Tukey-Kramer</th>
<th>P-I Pearson correlation coefficient</th>
<th>S-I Pearson correlation coefficient</th>
<th>P-S Pearson correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007 Group C</td>
<td>23</td>
<td>n.s. p=0.086</td>
<td></td>
<td>0.45</td>
<td>0.67</td>
</tr>
<tr>
<td>2006-2007 Group D</td>
<td>18</td>
<td>n.s. p=0.068</td>
<td></td>
<td></td>
<td>0.65</td>
</tr>
<tr>
<td>2007-2008 Group E</td>
<td>30</td>
<td>sig. p=0.0007 A=B&gt;C</td>
<td></td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>2007-2008 Group F</td>
<td>23</td>
<td>n.s. p=0.119</td>
<td>0.46</td>
<td>0.42</td>
<td>0.44</td>
</tr>
<tr>
<td>2008-2009 Group G</td>
<td>30</td>
<td>n.s. p=0.062</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009 Group H</td>
<td>19</td>
<td>n.s. p=0.384</td>
<td>0.52</td>
<td>0.45</td>
<td></td>
</tr>
</tbody>
</table>

n.s. = non-significant  sig. = significant  $A =$ Self  $B =$ Peer  $C =$ Instructor-researcher  $S-I =$ Self-Instructor  $P-I =$ Peer-Instructor  $P-S =$ Peer-Self

In the final year of the Main Study, 2008-2009, we find that the previous correlations between $P-S$ have disappeared and that there are now only correlations
between P-I in both groups G and H, with values of 0.44 and 0.52 respectively, while Group H, as in the case of Group F in the previous year, also has a correlation between S-I, this time with a value of 0.45. Again, this shows an encouraging tendency towards assessment alignment.

5.2.5 AARP Main Study Questionnaire results
Qualitative feedback concerning assessment processes was collected from AARP Main Study participants in the same way as described for participants in the Pre-Study, in Section 5.1.5, so that copies of Antonopoulou’s questionnaire (in the Greek version, see Appendix 21) were distributed to members of all six research groups. In addition, the same abridged format of Antonopoulou’s questionnaire (see Appendix 23) was used for purposes of AARP analysis.

In the year 2006-2007, 13 students in Group C and 16 students in Group D completed questionnaires, while in 2007-2008, 29 students in Group E and 21 students in Group F submitted them. In the final year of the Main Study, 2008-2009, 26 students in Group G and 20 in Group H completed questionnaires. Percentages for each point on the Likert scale of 1-5 for each of the ten questions for the 3 years of the Main Study are presented in the xyz graphs displayed in Appendices 27a to 27j in Vol. 2 of this thesis.

As in the case of the Pre-Study, the Main Study questionnaire data was also subjected to t-analysis of the means of each response (see Tables 5.16, 5.17 and 5.18 below) and shows complete agreement between all six groups from the Main Study in their responses to questions 1, 2, 3, 8, 9 and 10.

There is also agreement between Groups E, F, G and H on Questions 4 and 6 and between C, D, G and H on Question 7. This means that as in the Pre-Study with Groups A and B, there is disagreement between Groups C and D on Question 4,
concerned with objectivity in peer-assessment, producing a mean difference of 0.59 and \( p = 0.044 \), where 0.05 is considered significant. Anxiety about peer-assessment is expressed by members of both groups in the student comments in Appendix 27d.

**Table 5.16 Mean values of Group C and D (2006-2007) questionnaire responses and independent t-test results**

<table>
<thead>
<tr>
<th>Question</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>13</td>
<td>4.08</td>
<td>0.494</td>
<td>t(27)=1.299</td>
<td>0.205</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>4.31</td>
<td>0.479</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>13</td>
<td>4.00</td>
<td>0.707</td>
<td>t(27)=0.925</td>
<td>0.343</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>4.25</td>
<td>0.683</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>13</td>
<td>2.92</td>
<td>0.862</td>
<td>t(27)=1.643</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>3.38</td>
<td>0.619</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>13</td>
<td>3.54</td>
<td>0.776</td>
<td>t(27)=2.113</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>4.13</td>
<td>0.719</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>13</td>
<td>2.15</td>
<td>1.068</td>
<td>t(27)=2.194</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>2.94</td>
<td>0.854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>13</td>
<td>3.23</td>
<td>1.092</td>
<td>t(27)=2.170</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>4.00</td>
<td>0.816</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>13</td>
<td>4.15</td>
<td>0.801</td>
<td>t(27)=1.302</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>4.50</td>
<td>0.632</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>13</td>
<td>4.23</td>
<td>0.439</td>
<td>t(27)=0.916</td>
<td>0.368</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>4.00</td>
<td>0.816</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>13</td>
<td>4.54</td>
<td>0.519</td>
<td>t(27)=0.402</td>
<td>0.691</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>4.63</td>
<td>0.619</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>13</td>
<td>3.92</td>
<td>0.494</td>
<td>t(27)=1.873</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>16</td>
<td>4.31</td>
<td>0.602</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Likewise, for Question 5, which is concerned with how easy self-assessment is, for Groups C and D (see Table 5.16 above) there was a mean difference of 0.79 and a value of \( p=0.037 \), which is also significant. The difficulties of dealing with self-assessment are also expressed in the student comments in Appendix 27e. In the case of Question 6, there was a mean difference of 0.77 and a value of \( p=0.039 \), which is also significant. This indicates strong differences of opinion concerning peer- and self-assessment processes between the two groups in the year 2006-2007. A comment from a member of Group D, in Appendix 27f, encapsulates their difficulties with self-assessment.
In the following years, there was disagreement only over one question, which was Question 7 in 2007-2008 (see Table 5.17, below), concerned with the usefulness of the assessment criteria checklist in understanding areas in which learners were being assessed. Here there was a mean difference of 0.49 between Groups E and F and a value of \( p = 0.029 \). Only these two groups had disagreement on this particular question.

**Table 5.17 Mean values of Group E and F (2007-2008) questionnaire responses and independent t-test results**

<table>
<thead>
<tr>
<th>Question</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>E</td>
<td>23</td>
<td>3.74</td>
<td>0.449</td>
<td>t(42)=0.124</td>
<td>0.902</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>3.71</td>
<td>0.845</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td>23</td>
<td>3.61</td>
<td>0.656</td>
<td>t(42)=0.509</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>3.71</td>
<td>0.717</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>23</td>
<td>3.22</td>
<td>0.902</td>
<td>t(42)=1.142</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>3.52</td>
<td>0.873</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>23</td>
<td>3.91</td>
<td>0.793</td>
<td>t(42)=0.234</td>
<td>0.816</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>3.86</td>
<td>0.793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>23</td>
<td>3.17</td>
<td>0.717</td>
<td>t(42)=0.492</td>
<td>0.625</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>3.05</td>
<td>0.973</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>E</td>
<td>23</td>
<td>3.74</td>
<td>0.864</td>
<td>t(42)=0.243</td>
<td>0.809</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>3.67</td>
<td>1.111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>E</td>
<td>23</td>
<td>4.30</td>
<td>0.765</td>
<td>t(42)=2.261</td>
<td><strong>0.029</strong></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>3.81</td>
<td>0.680</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>E</td>
<td>23</td>
<td>3.91</td>
<td>0.848</td>
<td>t(42)=0.200</td>
<td>0.843</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>3.86</td>
<td>1.014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>E</td>
<td>23</td>
<td>4.35</td>
<td>0.982</td>
<td>t(42)=1.006</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>4.10</td>
<td>0.625</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>E</td>
<td>23</td>
<td>3.91</td>
<td>0.900</td>
<td>t(42)=1.144</td>
<td>0.259</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>21</td>
<td>4.19</td>
<td>0.680</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the final year of the Main Study, in Groups G and H (see Table 5.18 below), there was disagreement on one of the three questions which had also troubled Groups C and D, and this was Question 5, which concerned how easy they found Self-assessment. Here there was a mean difference of 0.83 and a value of \( p=0.010 \), which was significant. A comment from a member of Group H, in Appendix 27e, indicates that Self-assessment was considered to be as challenging as Peer-assessment.
Table 5.18  Mean values of Group G and H (2008-2009) questionnaire responses and independent t-test results

<table>
<thead>
<tr>
<th>Question</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>26</td>
<td>3.73</td>
<td>0.724</td>
<td>t(44)=0.571</td>
<td>0.571</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>3.85</td>
<td>0.671</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>26</td>
<td>4.00</td>
<td>0.800</td>
<td>t(44)=0.569</td>
<td>0.572</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>3.85</td>
<td>0.988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>26</td>
<td>3.58</td>
<td>0.703</td>
<td>t(44)=2.006</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>3.10</td>
<td>0.912</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>26</td>
<td>3.85</td>
<td>0.834</td>
<td>t(44)=0.411</td>
<td>0.683</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>3.75</td>
<td>0.716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>26</td>
<td>3.38</td>
<td>1.023</td>
<td>t(44)=2.692</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>2.55</td>
<td>1.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>26</td>
<td>3.81</td>
<td>0.939</td>
<td>t(44)=1.701</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>3.30</td>
<td>1.081</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>26</td>
<td>4.04</td>
<td>0.528</td>
<td>t(44)=1.040</td>
<td>0.304</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>3.80</td>
<td>1.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>G</td>
<td>26</td>
<td>3.96</td>
<td>0.720</td>
<td>t(44)=1.044</td>
<td>0.302</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>3.70</td>
<td>0.979</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>G</td>
<td>26</td>
<td>3.92</td>
<td>0.977</td>
<td>t(44)=0.254</td>
<td>0.801</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>4.00</td>
<td>1.076</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>G</td>
<td>26</td>
<td>3.77</td>
<td>0.863</td>
<td>t(44)=0.876</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>20</td>
<td>4.00</td>
<td>0.918</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More of the written responses to the questionnaire elicited from Group members in the Main Study offer further insights into their attitudes and beliefs and these can be viewed in the student comments sections in Vol.2, Appendices 27a – 27j.

5.2.6  AARP Main Study - Overview of outcome

The four research questions on which the AARP Main Study and the overall thesis are based are set out in Chapter 1, Section 1.4. While, in the AARP Pre-Study, the answers to the first three of the research questions could be answered fairly straightforwardly, when it comes to the Main Study, things are a little more complex.

In 5 out of the 6 research samples in the Main Study, the results for the assessment of Oral assignments are very similar to those of the Pre-Study. This means that the accuracy of peer-assessment and self-assessment of Oral skills as compared with the I-R is somewhere around 83%, which is still very admirable. It demonstrates
objectivity and reliability, it shows that the practice gained through peer-assessment is sufficient to be capable of self-assessment and it also reveals that learners take on ownership of the oral assessment criteria and use them in an atmosphere of cooperation and trust.

In the case of the assessment of 1st and 2nd Writing Assignments in the Main Study, things are not quite so transparent. If we judge the assessment of the 1st Writing Assignment from the Means of P-A, S-A and I-A, and ANOVA, then a 50% success rate could be claimed. If the correlation coefficients are also taken into account, then the success rate in the Main Study looks closer to 60%. If we judge assessment of the 2nd Writing Assignment in the same way, then things look much more disappointing and we would have to claim only 16.6% success based on Means and ANOVA. However, there are three positive correlations between Self and Instructor, ranging from 0.47 to 0.66, which certainly improve the picture significantly, showing that in one case, in 2006-2007, and in both cases in 2008-2009, participants in the AARP were more aware of the criteria and of their strengths and weaknesses and were trying hard to be honest with others and with themselves.

There is one more source of information concerning assessment of the 2nd Writing Assignment which has not yet been presented and that concerns the chi-square test (see Vol. 2, Appendix 19) from which rating frequency ratios were derived and these are presented in the fourth section of this chapter (see Section 5.4.2). What is apparent from Figure 5.2, in the same section, is the surprisingly close alignment, throughout the duration of the Main Study, which is revealed in Instructor:Peer (I:P) rating frequency ratios.

With regard to the 4th research question, we are forced to conclude that it was answered satisfactorily only in part. Students were willing to conduct honest and accurate assessment in the case of Oral skills, with the exception of Group E, but
resistance as regards Writing assessment, which existed initially in the larger groups, for reasons unknown, in the 1st Writing assessment cycle, spread also to the smaller groups in the 2nd Writing assessment cycle, with the exception of Group F. In the case of writing assessment and of responding to questionnaires, it seems that some learners simply “put on a mask of autonomous behaviour” (Breen & Mann, 1997, p. 141). Considering the results of the Preliminary Study, this was a somewhat disappointing outcome. The ups and downs and seeming deterioration in assessment behaviour, particularly with regard to Writing, over the three years of the Main Study, compared to the Pre-Study, prompted the I-R to resort to intervention in the Post-Study, in the fifth and final year of the AARP. The results of the Post-Study will be presented in the sections which follow.

5.3 AARP Post-Study Results

The AARP Post-Study was conducted in the last year of the AARP, which was 2009-2010, and was also conducted in the SOE, Auth. The participants in the study were once again students from 1st year Language Mastery I courses and the tasks, as in the Pre- and Main Studies, involved Self-, Peer- and Instructor assessment of 2 home writing assignments and 1 oral presentation given before a whole-class audience. Details of course procedures, tasks and assessment processes and research instruments have already been outlined in detail in Chapter 4.

Having established in the AARP Pre-Study in the academic year 2005-2006 that all procedures, tasks and research instruments were satisfactory, the same methods and methodology were replicated in the three years of the Main Study that followed, but this time produced different and quite mixed results. It was therefore deemed necessary in the final year of the AARP to attempt some assessment training
intervention techniques in order to see if results could be brought back to the levels of 2005-2006.

The aims of the AARP Post-Study were therefore twofold. The first aim, as in the previous years of the AARP, was essentially to find answers to the four research questions on which the hypothesis of the thesis is based, which were outlined in Chapter 1, Section 1.4.

The second aim of the AARP Post-Study was to determine if Intervention techniques, both in the assessment of Oral Skills and of Writing Skills, could perhaps bring assessment accuracy back to the apparently successful levels of the Pre-Study, 2005-2006. The techniques and processes used in the Intervention exercises have been described in Chapter 4, and here the results of the Intervention exercises will be presented, in addition to the results from the two regular LM I Writing assignments and the regular LM I Oral assignment.

5.3.1 Breakdown of Participants in the AARP Post-Study, 2009-2010
The research groups used in this AARP Post-Study on the course, LM I, in the autumn semester of 2009-2010, as in the Preliminary Study and the Main Study were the two groups of Language Mastery I, which will be termed I and J to preserve anonymity, which were assigned to the I-R on a random basis from a total of ten LM I groups. The breakdown of the two groups according to age, ethnic origin and language ability is shown below in Table 5.19.

As in the Pre- and Main Studies, individual participation in oral and writing tasks fluctuated in the Post-Study, as did attendance on the course, during the 5 years, whether compulsory or non-compulsory, but the maximum number of participants in any one task from Group I was 27, while the maximum number of participants from Group J was 22. In both groups, the female gender was predominant, as is typical of
Table 5.19  AARP Post-Study, 2009-2010 - Group breakdown

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>Group</th>
<th>Participants</th>
<th>Female</th>
<th>Male</th>
<th>17-19</th>
<th>20 and over</th>
<th>Greek/Cyp.</th>
<th>Other nat.</th>
<th>LEVEL AT ENTRY</th>
<th>LEVEL PREVIOUSLY ATTAINED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C2</td>
<td>C1</td>
<td>B2</td>
<td>B1</td>
<td>A2</td>
<td>U</td>
<td>C2</td>
<td>C1</td>
<td>B2</td>
</tr>
<tr>
<td>2009-2010</td>
<td>I</td>
<td>27</td>
<td>23</td>
<td>4</td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>17</td>
<td>5</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td>49</td>
<td>40</td>
<td>9</td>
<td>39</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

Levels shown are CEFR levels. U = Unknown
the School of English student population as a whole, with 23 females and 4 males in Group I and 17 females and 5 males in Group J. Likewise, the ethnic breakdown was predominantly Greek/Greek Cypriot, with 26 Greeks/Greek Cypriots and 1 of another nationality in Group I and 20 Greeks/Greek Cypriots and 2 of other nationalities in Group J. 21 of the participants in Group I were aged 17-19, while the remaining 6 were aged 20 or over, while in Group J, 18 of the participants were aged 17-19, with the remaining 4 aged 20 or over.

The level of English of the participating students, based on completion of the Oxford Placement Test (UCLES, 2001) revealed that in Group I, 2 were C2 level, 9 were C1 level, 9 were B2 level, 6 were B1 level, and none were A2 level. The level of the 1 remaining participant, due to late joining of the course and non-completion of the OPT, was unknown, but since this participant was repeating the course, their level was most likely B2 or below. In Group J, 2 were found to be C2 level, 6 were level C1, 7 were level B2, 4 were level B1, and none were level A2. Information was also collated about the students’ previous attainment in the language (see Table 5.19 above) and certificates gained, which was a useful indication of levels of 2nd language attrition, usually attributable to non-maintenance of the language during their last two years of Senior High School (Everhard, unpublished data).

5.3.2 AARP Post-Study Assessment results
In the sections which follow, the results of triangulated Self-, Peer- and Instructor assessment of the two LM I Home Writing Assignments as well as one LM I in-class Oral Assignment will be presented; however, because Intervention exercises were introduced for the 1st Writing Assignment as well as the Oral Assignment, the results of the Intervention exercises will be presented before the results of the assessment of the assignments in each case. It becomes clear that the reasonably good Peer-
assessment results produced in the mock Intervention exercises for these assignments is not always matched by the actual Peer- and Self-assessment processes for the same assignments. Non-participation of some students in the Intervention exercises appears to have resulted in distortion and ‘tainting’ of assessment data and therefore analysis of the data for each writing and oral assignment has been run twice: a first time with all assessment participants, including all non-Intervention exercise (IE) participants, and a second time with only bona fide IE participants. This second analysis permits a more accurate picture of whether the IE was, in fact, effective.

5.3.2.1 AARP Post-Study 1st Writing Assignment Intervention Exercise results
The reasons for intervention in the 1st Writing Assignment and a description of the processes involved can be found in Chapter 4, Section 4.6.1.1. In order to understand the results of the Intervention exercise (IE), the mean values for the participants in P-A from the two Groups I and J were calculated for each of the 5 writing samples and compared with the original S-A, P-A and I-A values that had been awarded. The values on which calculations were based are shown in Table 5.20 below:

| Table 5.20 AARP Post-Study - Values for 1st Writing Assessment Intervention |
|----------------------------------------|---|---|---|---|
| **Group I** 2009-2010 n=19            | S-A (A) | P-A (B1) | P-A (B2) | I-A (C ) |
| Writing Sample 1                       | 9.8     | 8.4      | 9.2      | 9.4      |
| Writing Sample 2                       | 6.8     | 7.4      | 6.4      | 6.8      |
| Writing Sample 3                       | 9.8     | 9.4      | 9.2      | 9.6      |
| Writing Sample 4                       | 6.6     | 7.6      | 7.6      | 7.8      |
| Writing Sample 5                       | 7.6     | 5.8      | 8.2      | 8.0      |
| **Group J** 2009-2010 n=19            | S-A (A) | P-A (B1) | P-A (B2) | I-A (C ) |
| Writing Sample 1                       | 9.8     | 8.4      | 9.0      | 9.4      |
| Writing Sample 2                       | 6.8     | 7.4      | 6.2      | 6.8      |
| Writing Sample 3                       | 9.8     | 9.4      | 8.4      | 9.6      |
The number of participants in the IE was 19 in both groups and the grades given above, with the exception of P-A (B2), are true scores out of 10 which were authentic ratings awarded to these samples by Self-assessors, Peer-assessors and the I-R in previous years and which were reached using the criteria checklists for the LM I 1st Writing Assignment (see Appendix 6).

By taking the true scores for each sample from each of the IE participants, an evaluation of the mock-assessment process was conducted using one-way analysis of variance (ANOVA) which produced values of $p=0.908$, in the case of Group I and $p=0.842$ in the case of Group J, indicating that this mock peer-assessment did not discriminate compared to the other S-A, P-A and I-A assessment values and therefore had no significant differences. This attempt at mock Peer Writing assessment could therefore be considered to have been successful for both groups and this is supported overall by student comments concerning training for Writing assessment (see Appendices 28d, 28e and 28i).

### 5.3.2.2 AARP Post-Study (Actual) 1st Writing Assignment assessment results

It should be mentioned that due to absences, negligence and lateness when the Intervention exercises were distributed, performed, discussed or collected in, not all students took part in the writing IE in the Post-Study and so some students were given Peer-assessment to do in the usual way at a later date, while all students at a later date were given their Self-assessment tasks to complete in the usual way. The I-R completed the final stage of the assessment cycle with Instructor-assessment and the

<table>
<thead>
<tr>
<th>Writing Sample 4</th>
<th>6.6</th>
<th>7.6</th>
<th>7.3</th>
<th>7.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Sample 5</td>
<td>7.6</td>
<td>5.8</td>
<td>7.6</td>
<td>8.0</td>
</tr>
</tbody>
</table>

S-A (A) = Actual Self-assessment  
P-A (B1) = Actual Peer-assessment  
P-A (B2) = Means of Mock Peer-assessment  
I-A (C) = Actual Instructor Assessment
grades from P-A, S-A and I-A were then recorded on the Student Profile cards (see Appendix 3), as well as electronically.

The grades produced from the total S-A, P-A and I-A triangulation process, including both participants and non-participants in the Intervention exercises, were subjected to statistical analysis and produced the following, as shown in Table 5.21 below. The values calculated only with the 18 students who took part in the IE have been placed in brackets and are given below the regular (Actual) mean score in each case, for comparison:

### Table 5.21 AARP Post-Study 1st Writing Assignment Mean Scores and ANOVA results

<table>
<thead>
<tr>
<th></th>
<th>2009-2010</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A (A)</td>
<td>8.80</td>
<td>0.79</td>
<td>0.149</td>
<td>28</td>
<td>F(2,81)=8.981</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.88)</td>
<td>(0.85)</td>
<td>(0.201)</td>
<td>(18)</td>
<td>(F(2,51)=9.157)</td>
<td>(0.0004)</td>
<td></td>
</tr>
<tr>
<td>P-A (B)</td>
<td>8.13</td>
<td>0.99</td>
<td>0.187</td>
<td>28</td>
<td>F(2,81)=8.981</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.02)</td>
<td>(1.05)</td>
<td>(0.248)</td>
<td>(18)</td>
<td>(F(2,51)=9.157)</td>
<td>(0.0004)</td>
<td></td>
</tr>
<tr>
<td>I-A (C)</td>
<td>7.21</td>
<td>1.28</td>
<td>0.242</td>
<td>28</td>
<td>F(2,81)=8.981</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.06)</td>
<td>(1.38)</td>
<td>(0.325)</td>
<td>(18)</td>
<td>(F(2,51)=9.157)</td>
<td>(0.0004)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2009-2010</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>N</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group J</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-A (A)</td>
<td>8.30</td>
<td>0.98</td>
<td>0.205</td>
<td>23</td>
<td>F(2,66)=2.850</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.41)</td>
<td>(1.01)</td>
<td>(0.237)</td>
<td>(18)</td>
<td>(F(2,51)=3.072)</td>
<td>(0.055)</td>
<td></td>
</tr>
<tr>
<td>P-A (B)</td>
<td>7.96</td>
<td>1.18</td>
<td>0.245</td>
<td>23</td>
<td>F(2,66)=2.850</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.07)</td>
<td>(1.08)</td>
<td>(0.254)</td>
<td>(18)</td>
<td>(F(2,51)=3.072)</td>
<td>(0.055)</td>
<td></td>
</tr>
<tr>
<td>I-A (C)</td>
<td>7.52</td>
<td>1.14</td>
<td>0.237</td>
<td>23</td>
<td>F(2,66)=2.850</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.54)</td>
<td>(1.08)</td>
<td>(0.255)</td>
<td>(18)</td>
<td>(F(2,51)=3.072)</td>
<td>(0.055)</td>
<td></td>
</tr>
</tbody>
</table>

S-A = Self-assessment   P-A = Peer-assessment   I-A = Instructor assessment

In both Groups I and J, the highest mean grade was awarded by S-A, with values of 8.81 (8.88) and 8.30 (8.41) respectively, followed in descending order by P-A, with values of 8.13 (8.02) and 7.96 (8.07) respectively, and finally a more severe I-A, with a much lower value of 7.21 (7.06), creating a Mean Difference of 0.92 (0.97) with P-A, in Group I and a more moderate I-A value of 7.52 (7.54) in Group J, creating a smaller Mean Difference with P-A, of 0.44 (0.52). The highest level of Standard Deviation (SD) occurred in I-A, in Group I, with a value of 1.28 (1.38), followed by P-A with 0.99 (1.05), indicating that the I-R was using a wider range of ratings, while this situation is reversed in Group J, with the highest SD occurring in P-A, with a
value of 1.18 (1.08), followed by I-A, with 1.14 (1.08)\textsuperscript{18}, indicating that Peer-assessors were awarding a wider range of ratings.

One-way ANOVA revealed a \( p \) value of 0.0003 (0.0004)\textsuperscript{19} for Group I, which is considered significant, while the \( p \) value for Group J was found to be 0.065 (0.055), which is considered not significant. A Tukey-Kramer Multiple Comparison Test of Group I revealed a pattern of A>B>C\textsuperscript{20}, this being a pattern which had never occurred for this assignment previously. Likewise, analysis of Pairwise Relationships revealed a statistically significant correlation between P-A and I-A in Group I, with a correlation coefficient of 0.42 and in Group J, with a coefficient of 0.62\textsuperscript{21} (see Table 5.22 below), which was an encouraging sign of alignment.

<table>
<thead>
<tr>
<th>( p = 0.05 )</th>
<th>n</th>
<th>One-way ANOVA &amp; Tukey-Kramer</th>
<th>P-I Pearson correlation coefficient</th>
<th>S-I Pearson correlation coefficient</th>
<th>P-S Pearson correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-2010 Group I</td>
<td>28</td>
<td>sig. ( p &lt; 0.001 ) A&gt;B&gt;C</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010 Group J</td>
<td>18</td>
<td>sig. ( p &lt; 0.001 ) A&gt;B&gt;C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010 Group J</td>
<td>23</td>
<td>n.s.</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010 Group J</td>
<td>18</td>
<td>n.s.</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n.s. = non-significant, sig. = significant

A = Self
B = Peer
C = Instructor-researcher
P-I = Peer-Instructor
P-S = Peer-Self
S-I = Self-Instructor

On completion of the assessment cycle, the three copies of the assignment and their assessments were returned to the owners and discussion of the assessment outcomes

\textsuperscript{18} The Standard Deviation values for P-A and I-A when we include only the trained students is almost equal, with 1.078 for P-A and 1.080 for I-A, indicating that they were rating within the same range of grades.

\textsuperscript{19} The difference in the \( p \) value when including only the 18 trained participants or including the 10 untrained participants is only slight.

\textsuperscript{20} The pattern changes to A=B>C.

\textsuperscript{21} The correlation coefficient in Group I disappears when only the trained participants are included, while it rises to 0.66 in Group J.

\textsuperscript{22} These are the results when only students who participated in the Intervention for the 1\textsuperscript{st} Writing Assignment are included.

\textsuperscript{23} These are the results when only students who participated in the Intervention for the 1\textsuperscript{st} Writing Assignment are included.
took place in meetings with the I-R, in her office or in the RC, so that further clarification and feedback could be provided by the I-R.

5.3.3 AARP Post-Study 2nd Home Writing Assignment assessment results

Since intervention in the 1st Home Writing Assignment had added even further to the I-R’s workload and did not seem to have produced any radical changes in assessment behaviour or any startlingly surprising results, the I-R decided to abandon any further intervention in writing assessment preparation and to carry out procedures for the 2nd Home Writing Assignment and its assessment in the same way as in the Pre-Study and years 2007-2008 and 2008-2009 in the Main Study. The assignment was set and collected in the same manner and the scoring from Peer-assessment, Self-assessment and Instructor-Assessment was collected and collated in the same way, with records being kept, both on Student Profile Cards (see Appendix 3) and electronically.

When the grades had been recorded, they were subjected to statistical analysis, which revealed the following, as shown in Table 5.23 below. The values, when calculated to include only those who took part in the 1st Writing Assignment assessment IE are given below the Actual, in each case, for comparison:

### Table 5.23 AARP Post-Study 2nd Writing Assignment Mean Scores and ANOVA results

<table>
<thead>
<tr>
<th>2009-2010 Group I</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A (A)</td>
<td>8.84</td>
<td>0.80</td>
<td>0.154</td>
<td>27</td>
<td>F(2,78)=7.557</td>
<td>0.001</td>
</tr>
<tr>
<td>P-A (B)</td>
<td>7.79</td>
<td>1.11</td>
<td>0.213</td>
<td>27</td>
<td>F(2,51)=4.171</td>
<td>0.021</td>
</tr>
<tr>
<td>I-A (C)</td>
<td>8.53</td>
<td>1.11</td>
<td>0.214</td>
<td>27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2009-2010 Group J</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A (A)</td>
<td>8.59</td>
<td>0.95</td>
<td>0.199</td>
<td>23</td>
<td>F(2,66)=1.463</td>
<td>0.239</td>
</tr>
<tr>
<td>P-A (B)</td>
<td>8.35</td>
<td>1.05</td>
<td>0.220</td>
<td>23</td>
<td>F(2,51)=1.517</td>
<td>0.229</td>
</tr>
<tr>
<td>I-A (C)</td>
<td>8.02</td>
<td>1.38</td>
<td>0.288</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S-A = Self-assessment    P-A = Peer-assessment    I-A = Instructor assessment
Although, as previously mentioned, there was no intervention on this occasion, it should be noted that there was a significant change in the assessment behaviour of Group I. In both Group I and Group J, the highest mean value was produced by S-A, with a value of 8.84 (8.83) and 8.59 (8.68) respectively, but in the case of Group I, the next-ranking value was produced by I-A, with a value of 8.53 (8.69), as opposed to the next-ranking P-A value of 8.35 (8.33) in Group J. Thus, the lowest mean value was produced by P-A in Group I, with a value of 7.79 (8.04), and by I-A in Group J, with a value of 8.02 (8.06). In Group I, both P-A and I-A produced very similar Standard Deviation levels of 1.11 (1.01) and 1.11 (0.91) respectively, indicating that they were rating within a similar range, while in Group J, I-A produced the highest level of Standard Deviation with a value of 1.38 (1.20), indicating that the I-R was awarding a wider range of ratings.

One-way ANOVA produced a $p$ value of 0.001 (0.021) in the case of Group I, which is considered significant, and produced a $p$ value of 0.239 (0.229) in the case of Group J, which is considered not significant. A Tukey-Kramer Multiple Comparison Test of Group I revealed a pattern of $A=C>B$ ($A>C$ only), while analysis of Pairwise Relationships revealed a statistically significant correlation between I-A and P-A, with a correlation coefficient of 0.52 (see Table 5.24 below).

---

24 The $p$ value for Group I changes to 0.021 with only the 18 trained participants, which is significant, while the pattern changes to $A>C$ only. In the case of Group J, the $p$ value is 0.229, which is considered not significant and all means are equal.

25 This correlation between P-A and I-A rises to a coefficient value of 0.65, when only the 18 trained participants are included.
Table 5.24 AARP Post-Study 2nd Writing Assignment assessment – Tukey-Kramer mean comparisons and Pearson correlation coefficients – ‘pure’ and ‘tainted’ groups compared

<table>
<thead>
<tr>
<th>p=0.05</th>
<th>n</th>
<th>One-way ANOVA &amp; Tukey-Kramer</th>
<th>P-I Pearson correlation coefficient</th>
<th>S-I Pearson correlation coefficient</th>
<th>P-S Pearson correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-2010 Group I</td>
<td>27</td>
<td>sig. p=0.001 A=C&gt;B</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010 Group J</td>
<td>23</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010 Group J</td>
<td>18</td>
<td>sig. p=0.021 A&gt;C only</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010 Group J</td>
<td>18</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n.s. = non-significant sig. = significant A = Self B = Peer C = Instructor-researcher S-I = Self-Instructor P-I = Peer-Instructor P-S = Peer-Self

The change in the behaviour of P-A in Group I and the greater severity with which they rated as compared with the 1st Assignment is puzzling. One possible explanation could be that the assessment training Intervention which took place with the 1st Writing Assignment could have had a delayed effect, for reasons hard to explain, and does not show until the 2nd Writing Assignment. Cheng & Warren (1999) note a similar phenomenon, of what seems to be delayed effects, with their students, in the case of assessment of a 2nd Oral assignment, while Chen (2006b, p. 6) also reports “improvement of assessment accuracy” in a 2nd cycle of learner-centred Oral assessment.

As a result, we see closer convergence of rating between S-A and I-A, which is even closer when calculations include only those students who had taken part in the writing assessment IEs.

In the section which follows, our focus will turn from the assessment of Writing skills to the assessment of Oral skills.

26 These are the results when only students who participated in the Intervention for the 1st Writing Assignment are included.
27 These are the results when only students who participated in the Intervention for the 1st Writing Assignment are included.
5.3.4 AARP Post-Study Oral Assignment assessment and Intervention

The way in which procedures for the Oral assignment varied in the Post-Study, 2009-2010, have been outlined in Chapter 4, Section 4.6.3.1 and require no further elaboration. The outcome of the Oral Invention exercise (IE) itself, which resulted from assessment training with mock presentations, will be presented in the section which follows and therefore precedes the results of the Post-Study Oral Assignment assessment which are given in Section 5.3.4.2.

5.3.4.1 AARP Post-Study Oral Assessment Intervention exercise (IE) results

20 students from Group I and 17 from Group J took part in the Oral IE. In order to evaluate the outcome of the Oral IE in the Post-Study, the data from the 9 Mock Presentations was collated and analysed as shown in Table 5.25 below:

<table>
<thead>
<tr>
<th>Group I 2009-2010 n=20</th>
<th>P-A</th>
<th>95% CI_L</th>
<th>95% CI_U</th>
<th>I-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Mock Presentation</td>
<td>7.6</td>
<td>7.2</td>
<td>8.1</td>
<td>6.8</td>
</tr>
<tr>
<td>2nd Mock Presentation</td>
<td>9.3</td>
<td>9.1</td>
<td>9.6</td>
<td>8.8</td>
</tr>
<tr>
<td>3rd Mock Presentation</td>
<td>9.8</td>
<td>9.6</td>
<td>9.9</td>
<td>10</td>
</tr>
<tr>
<td>4th Mock Presentation</td>
<td>9.2</td>
<td>8.8</td>
<td>9.6</td>
<td>8.4</td>
</tr>
<tr>
<td>5th Mock Presentation</td>
<td>8.5</td>
<td>8.0</td>
<td>9.0</td>
<td>8.0</td>
</tr>
<tr>
<td>6th Mock Presentation</td>
<td>8.7</td>
<td>8.2</td>
<td>9.1</td>
<td>8.0</td>
</tr>
<tr>
<td>7th Mock Presentation</td>
<td>9.4</td>
<td>9.0</td>
<td>9.9</td>
<td>10</td>
</tr>
<tr>
<td>8th Mock Presentation</td>
<td>7.8</td>
<td>7.2</td>
<td>8.4</td>
<td>7.2</td>
</tr>
<tr>
<td>9th Mock Presentation</td>
<td>7.5</td>
<td>6.9</td>
<td>8.1</td>
<td>6.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group J 2009-2010 n=17</th>
<th>P-A</th>
<th>95% CI_L</th>
<th>95% CI_U</th>
<th>I-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Mock Presentation</td>
<td>6.7</td>
<td>6.1</td>
<td>7.2</td>
<td>6.8</td>
</tr>
<tr>
<td>2nd Mock Presentation</td>
<td>9.1</td>
<td>9.1</td>
<td>9.3</td>
<td>8.8</td>
</tr>
<tr>
<td>3rd Mock Presentation</td>
<td>9.5</td>
<td>9.5</td>
<td>9.8</td>
<td>10</td>
</tr>
</tbody>
</table>
For Group I, mean grades, which were scored out of a possible 10 using the assessment criteria check-lists for LM I Oral Presentations (see Appendix 10) were extracted for the 9 mock presentations from each of the 20 students who participated in the Oral IE. As can be determined from Table 5.25, P-A exceeded I-A in seven cases, while I-A exceeded P-A in two cases. In one of the nine cases, in the 3rd Mock Presentation, P-A and I-A came close to coinciding. In general, Group I is over-generous in its rating, to a significant degree.

There was greater alignment in grading in the case of Group J, where P-A exceeded I-A in four cases and I-A exceeded P-A in four cases and in one case P-A and I-A coincided; however, in three out of eight cases, agreement was very close indeed. This shows much closer alignment between learner-assessment and instructor-assessment in this group, since the learners have a tendency to mark more strictly in comparison with Group I. Some sample comments reflecting students’ attitudes towards the Oral IE can be found in Appendix 28q.

### 5.3.4.2 AARP Post-Study Oral Assignment assessment results

From thereon, the Peer- and Self-assessment process for Groups I and J was conducted in the usual way (see Chapter 4, Sections 4.5.2 and 4.6.3), with the usual assessment criteria checklists (see Appendix 10), again using lists of names with the assessment check-list forming a cover for the assessment booklet. Oral presentations
and their assessment could either be conducted as part of a regular lesson or on a special day assigned to presentations. The booklets were collected in for collation and grades were recorded on the Student Profile Cards and also electronically. Once recorded, the grades were subjected to statistical analysis, as shown in Table 5.26:

A comparison of the means of S-A, P-A and I-A revealed a similar pattern between the two groups, with the highest means occurring among S-A, followed in descending order by P-A and I-A. The S-A mean value for Group I was, rather surprisingly, the highest value ever recorded in the course of 5 years, with 9.22 as compared with 8.67 for Group J; however, the P-A value in Group I was equally inflated, with a value of 8.93, as opposed to 8.49 in Group J. The Mean for I-A was well behind S-A and P-A in Group I and I-A had a Standard Deviation of 1.27, while the I-A Mean for Group J was 8.42, which followed very close behind the P-A value of 8.49. The Mean Difference between S-A and I-A was 0.92 in the case of Group I, while it was only 0.26 in the case of Group J, the latter showing impressive alignment between self-raters and the I-R.

One-way ANOVA revealed a p value of 0.001 in the case of Group I, which is considered significant, while the p value of 0.563, in the case of Group J is considered

<table>
<thead>
<tr>
<th>2009-2010 Group I</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A (A)</td>
<td>9.22</td>
<td>0.83</td>
<td>0.159</td>
<td>27</td>
<td>F(2,78)=2.835</td>
<td>0.001</td>
</tr>
<tr>
<td>P-A (B)</td>
<td>8.93</td>
<td>0.46</td>
<td>0.089</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A (C)</td>
<td>8.30</td>
<td>1.27</td>
<td>0.244</td>
<td>27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2009-2010 Group J</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A (A)</td>
<td>8.67</td>
<td>0.77</td>
<td>0.163</td>
<td>22</td>
<td>F(2,63)=0.580</td>
<td>0.563</td>
</tr>
<tr>
<td>P-A (B)</td>
<td>8.49</td>
<td>0.47</td>
<td>0.100</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-A (C)</td>
<td>8.42</td>
<td>1.00</td>
<td>0.213</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S-A = Self-assessment  P-A = Peer-assessment  I-A = Instructor assessment
not to be significant. A Tukey-Kramer Multiple Comparison Test of Group I revealed a pattern of A=B>C, while analysis of Pairwise Relationships revealed no statistically significant correlation coefficients.

Because of this strange turn of events with regard to oral assessment in the 5th Year of the AARP, and the resulting significant differences in Group I, which was something exceptional which had only once before occurred in oral assessment, with Group E in 2007-2008, the I-R decided that further investigation was required. It was decided to run the ANOVA tests once again, but this time including only the students who had participated in the Oral Intervention exercise (IE). This meant that only 20 of the 27 participants from Group I would be included this time and only 18 of the 22 participants from Group J. The results of the re-analysis are presented in Table 5.27 below:

Table 5.27  AARP Post-Study Oral Assignment Mean Scores and ANOVA results with Intervention participants only

<table>
<thead>
<tr>
<th>2009-2010 Group I</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A (A) 9.10</td>
<td>0.87</td>
<td>0.194</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-A (B) 8.86</td>
<td>0.53</td>
<td>0.118</td>
<td>20</td>
<td></td>
<td>F(2,57)=2.957</td>
<td>0.060</td>
</tr>
<tr>
<td>I-A (C) 8.40</td>
<td>1.25</td>
<td>0.278</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2009-2010 Group J</th>
<th>Mean</th>
<th>SD</th>
<th>SEM</th>
<th>n</th>
<th>F-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-A (A) 8.56</td>
<td>0.78</td>
<td>0.183</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-A (B) 8.52</td>
<td>0.45</td>
<td>0.106</td>
<td>18</td>
<td></td>
<td>F(2,51)=0.025</td>
<td>0.975</td>
</tr>
<tr>
<td>I-A (C) 8.58</td>
<td>0.97</td>
<td>0.229</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S-A = Self-assessment    P-A = Peer-assessment    I-A = Instructor assessment

These results are very interesting indeed. In Group I, the means for S-A maintain the highest position, but the mean difference between S-A and I-A has narrowed from 0.919 to a difference of 0.70. What is even more remarkable is that in Group J, I-A has overtaken S-A and has the highest ranking means, with 8.58, as compared with an S-A means of 8.56, and P-A follows close behind with 8.52. Alignment between the three categories of S-A, P-A and I-A appears to be very close indeed.
With regard to Standard Deviation (SD), all of them have dropped slightly, but I-A retains the highest levels, with 1.25 and 0.97 in Groups I and J respectively, indicating that the I-R is rating within a wider range, while S-A SD levels follow, with 0.87 and 0.78 respectively, whereas SD levels from P-A are much lower, at levels of 0.53 and 0.45 respectively, revealing a much narrower range of marking amongst Peers. One-way Analysis of Variance (ANOVA) revealed a \( p \) value of 0.06 in the case of Group I and a \( p \) value of 0.975, in Group J, both of which are considered non-significant. Further statistical analysis of oral assessment data, using Pairwise Relationships revealed a correlation coefficient for Group I of 0.51 between P-A and I-A (P-I), where \( p = 0.021 \), indicating assessment alignment between Peers and the I-R. The two possible interpretations of the oral assessment data are highlighted in Table 5.28 below:

**Table 5.28 AARP Post-Study Oral assignment assessment results – Tukey-Kramer mean comparisons and Pearson correlation coefficients – Mixed, and Intervention participants only**

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>One-way ANOVA &amp; Tukey-Kramer</th>
<th>P-I Pearson correlation coefficient</th>
<th>S-I Pearson correlation coefficient</th>
<th>P-S Pearson correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>27</td>
<td>sig. ( p=0.001 ) A=B&gt;C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>20</td>
<td>n.s.</td>
<td></td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Group J</td>
<td>22</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group J</td>
<td>18</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n.s. = non-significant  sig. = significant  A = Self  S-I = Self-Instructor  B = Peer  P-I = Peer-Instructor  C = Instructor-researcher  P-S = Peer-Self

This contrast in the comparison between the two sets of oral assessment data for Groups I and J seems to suggest that students who undertook the Oral Assessment IE

---

28 These are the ‘pure’ results if only those students who participated in the Intervention exercises are included in the data. Clearly, the 7 non-participants acted as outliers.

29 These are the ‘pure’ results if only those students who participated in the Intervention exercises are included in the data. Assessment behaviour is similar in both cases as can be seen in the Mean Scores in Tables 5.26 and 5.27 above.
exercised much greater control and restraint in their assessment practices. The results in both cases for Group J converge quite closely and are pleasing, while in Group I, distortion occurs in the first set of data and both S-A and P-A seem to have swung slightly out of control, which may be attributable to “cheating” (Dickinson, 1987, p. 150) or “friendship marking” (Sluijsmans et al., 1999, p. 302) among those seven students who did not take part in the Intervention. Looked at this way, it could be argued that the Oral assessment IE did, after all, bring success and had a very positive effect on those who took part. Some samples of student attitudes towards Oral assignment processes in the Post-Study can be read in Appendices 28n to 28v.

5.3.5 AARP Post-Study Intervention Questionnaire results
As described in Chapter 4, Section 4.3.2.2, a new questionnaire (see Appendix 25), compiled by the I-R, in English, was implemented at the end of the autumn semester in the Post-Study year, 2009-2010. The questionnaires were completed anonymously by 29 respondents, in the case of Group I, and 22 respondents, in the case of Group J, and the results of their responses to the full questionnaire in the Post-Study can be viewed in Vol. 2, Appendices 28a – 28v, where a commentary on the results is provided, together with xyz graphs, showing the percentage responses of students in both groups. By way of more qualitative data, some comments made by students, which provide further enlightenment as to their engagement in assessment and intervention processes, have been included in each case. These were given in English by the students and they have not been corrected or altered in any way.

To provide a more synoptical view of student responses to the questionnaire, Table 5.29 below shows the results of t-analysis of the mean scores of the responses to each question for both Groups I and J.
Table 5.29  Mean values of Group I and J (2009-2010) questionnaire responses and independent t-test results

<table>
<thead>
<tr>
<th>Question</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>4.07</td>
<td>0.842</td>
<td>t(49)=0.949</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>4.27</td>
<td>0.631</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.79</td>
<td>0.978</td>
<td>t(49)=1.895</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.23</td>
<td>1.152</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>4.10</td>
<td>1.113</td>
<td>t(49)=1.957</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.50</td>
<td>1.058</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>4.17</td>
<td>0.759</td>
<td>t(49)=0.777</td>
<td>0.441</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>4.00</td>
<td>0.816</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.93</td>
<td>0.923</td>
<td>t(49)=2.448</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.27</td>
<td>0.985</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.90</td>
<td>0.900</td>
<td>t(49)=0.689</td>
<td>0.494</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.73</td>
<td>0.827</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>4.34</td>
<td>0.721</td>
<td>t(49)=0.826</td>
<td>0.413</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>4.14</td>
<td>1.082</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.03</td>
<td>1.180</td>
<td>t(49)=0.741</td>
<td>0.462</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.27</td>
<td>1.077</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.34</td>
<td>1.111</td>
<td>t(49)=0.513</td>
<td>0.610</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.18</td>
<td>1.140</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.38</td>
<td>1.147</td>
<td>t(49)=0.054</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.36</td>
<td>0.848</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.72</td>
<td>1.131</td>
<td>t(49)=1.095</td>
<td>0.279</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>4.05</td>
<td>0.899</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>4.07</td>
<td>0.842</td>
<td>t(49)=1.019</td>
<td>0.313</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.82</td>
<td>0.907</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.90</td>
<td>0.860</td>
<td>t(49)=0.332</td>
<td>0.741</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.82</td>
<td>0.795</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.34</td>
<td>1.143</td>
<td>t(49)=1.279</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.73</td>
<td>0.935</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>4.17</td>
<td>0.805</td>
<td>t(49)=0.886</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.95</td>
<td>0.950</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.28</td>
<td>1.306</td>
<td>t(49)=0.897</td>
<td>0.374</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>2.95</td>
<td>1.214</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>4.45</td>
<td>0.827</td>
<td>t(49)=2.815</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.77</td>
<td>0.869</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.41</td>
<td>1.211</td>
<td>t(49)=1.210</td>
<td>0.232</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.82</td>
<td>1.140</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>4.41</td>
<td>0.780</td>
<td>t(49)=1.846</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.95</td>
<td>0.999</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.55</td>
<td>1.088</td>
<td>t(49)=1.215</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.91</td>
<td>0.971</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.69</td>
<td>1.137</td>
<td>t(49)=0.290</td>
<td>0.773</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>22</td>
<td>3.77</td>
<td>0.813</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>29</td>
<td>3.52</td>
<td>1.122</td>
<td>t(49)=0.256</td>
<td>0.799</td>
</tr>
</tbody>
</table>
As can be deduced, from Table 5.29 above, Group I was slightly more positive in its responses as compared with Group J and produced higher mean scores on 13 out of the 22 questions. There was, however, strong disagreement only on two questions: on Question 5, which was related to the need for training for writing assessment, in which Group I was more in favour of training, a further indication of this being given in student comments (see Appendix 28e), and on Question 17, which was concerned with the usefulness of the mock presentations as preparation for oral assessment, where again Group I was more strongly in favour of training through mock presentations. Rather mixed comments offered by the students on this subject can be found in Appendix 28q. It could be that Group J felt themselves capable of assessing without further training and this notion is, indeed, supported by their assessment behaviour.

5.3.6 AARP Post-Study - Overview of outcome
The use of Intervention, by its nature, in the AARP Post-Study, actually makes the answers to the research questions a much more complicated, but interesting, business. This is because, in effect, we are obliged to interpret three sets of data, rather than one. In the first set, the raw data has been treated in the exact same way as in previous years, but this then ignores the fact that not all students took part in the Intervention exercises, which means this data is, in a sense, distorted. The second set of data concerns the actual Intervention itself, while the third set of data takes the first set of data and filters out those who did not take part. It could be argued that only the second and third sets of data give a truly ‘pure’ picture of the Post-Study, the whole point of which was to determine the effects of Intervention. What this means is that we have two sets of answers to the research questions for the Post-Study, by this meaning that we could answer the research questions with recourse to the ‘pure’ data
and/or the ‘tainted’ data, with the latter including students who did not take part in Intervention exercises.

With regard to assessment of writing skills, if we judge only from the results of the writing assessment Intervention, which involves a type of peer-assessment only, then the answers to the first three research questions (see Chapter 1, Section 1.4) are all positive, with clear indications of objectivity and reliability, ownership of the criteria, and evidence of cooperation and trust. Some proof of this is provided in the Anonymous Sample of one student’s IE assessment (see Appendices 14a & 14b).

Concerning the results for the Actual assessment of the 1st and 2nd Writing Assignment in the Post-Study, the pattern is very reminiscent of Groups E and F in the year 2007-2008 of the Main Study, whether we look at the ‘pure’ or the ‘tainted’ results, with the first and larger group, Group I, showing less objectivity and reliability in its assessment accuracy than the second, Group J, but in both Post-Study groups there are indications of correlation and alignment between Peer-assessors and the I-R, particularly in the case of Group I, and this alignment strengthens in the 2nd Writing Assignment, which is a clear sign of growing objectivity and reliability as well as increasing cooperation and trust.

With regard to the assessment of the Oral Assignment in the Post-Study, again if we answer the first three research questions based on the results of the Intervention exercise, which is concerned purely with peer-assessment, we would be obliged to answer positively for Group J, and a little less positively for Group I, that the first and the third research questions are answered satisfactorily. Some evidence of this is provided in the anonymous Student Sample of participation in the Oral IE, where there are strong signs of alignment with the I-R’s assessment (see Appendix 16).

The great similarity with Groups E and F from the Main Study continues also with the assessment of Oral Skills, if we consider only the ‘tainted’ data for Groups I
and J in the Post-Study. If, on the other hand, we look at the ‘pure’ data, for Groups I and J, we are pleasantly reminded of all the other pairs of groups in the AARP, with Groups E and F excepted. In fact, looked at this way, the three research questions are answered positively by all groups in the AARP, with the exception of ‘rogue’ Group E, and Group I changes from a ‘rogue’ group to one which produces accurate peer- and self-assessment.

With regard to the 4th research question, it becomes clear from a reading of both the quantitative and qualitative data provided by the assessment questionnaires in the Post-Study that although not all students were completely happy with what for them seemed like very unconventional assessment processes, that they were nevertheless forced to think and deliberate and therefore engage with the written and spoken language at more than just a superficial level. The Intervention exercises seemed to have a positive effect on those who participated, albeit with a delay. The cyclical process of peer-, self- and instructor assessment seems to have been a learning opportunity for the majority of students, encouraging them to be more active learners, with a greater sense of empowerment and self-direction. The 4th research question would seem to be answered positively by the evidence gleaned from those involved in Intervention, but some degree of resistance is still evident from others.

One thing that must be said about the Post-Study results is that Intervention did not seem to make any particular impact on assessment behaviour with regard to writing, and may actually have been redundant in the case of Group J. With regard to speaking, it may have had some impact on those from Group I who took part, but may have again been redundant in the case of Group J. Although not all students took part in the Intervention exercises, it seems unlikely that total involvement would have made a difference and have helped to achieve the I-R’s goal of replicating the results of the AARP Pre-Study.
5.4 Rating frequency ratios throughout the AARP (2005-2010)

One final statistical analysis which can provide further information and insights about the assessment behaviour of the participants in the AARP, are rating frequency ratios derived from chi-square test analyses. Rating frequency ratios are useful in that they can help us understand the overall rating patterns and rating relationships between the Instructor and Self-assessors (I-S), between the Instructor and Peer-assessors (P-S) and between Peer-assessors and Self-assessors (P-S). Such relationships, between I-S, I-P and P-S will be examined in relation to assessment of the 1st Home Writing Assignment, the 2nd Home Writing Assignment, as well as the Oral Assignment on the AARP.

In the case of I-S and I-P, chi-analysis was used to calculate rating frequency ratios, based on:

(a) the number of scores of Self/Peer greater than the Instructor’s;
(b) the number of scores of Self/Peer lower than the Instructor’s;
(c) the number of scores that coincided with those of the Instructor.

In the case of P-S, calculations of rating frequency ratios were based on:

(a) the number of Peer scores greater than Self scores;
(b) the number of Peer scores lower than Self scores;
(c) the number of Peer scores that coincided with Self scores.

Based on these analyses, the frequency ratios for rating between I-S are expressed as I:S, between I-P are expressed as I:P and between P-S are expressed as P:S, and thus when we have a ratio, in any of these rating relationships, of 1:1, this is an indication of equalisation and perfect equilibrium.

The chi-square analyses used to calculate the rating frequency ratios can be found in Vol. 2 of this thesis, in Appendix 18, for the 1st Home Writing assignment, in Appendix 19, for the 2nd Home Writing assignment, and in Appendix 20, for the
Oral assignment. A point that should be noted in relation to the way that chi-square analyses were conducted, compared with other statistical analyses, is that while it has been very clear that individual groups within the pairs of research groups each year in the AARP were heterogeneous in nature, as regards their assessment behaviour, and have therefore been subjected to statistical analysis individually, for the purposes of chi-square testing, which requires a reasonable number of participants in order to offer reliability, the data from each annual pair of groups was pooled. Given below in Tables 5.30, 5.31 and 5.32 are the rating frequency ratios for the duration of the AARP, derived from the chi-square tests, which can be seen in Appendices 18, 19 & 20.

5.4.1 Rating frequency ratios for the AARP 1st writing assignment, 2005-2010
Table 5.30 below reveals the variations in Instructor-Self (I-S), Instructor-Peer (I-P) and Peer-Self (P-S) rating behaviour in assessment of the 1st Writing Assignment and how it changes and evolves over the five-year period of the AARP, between 2005-2010.

<table>
<thead>
<tr>
<th>Groups &amp; year</th>
<th>I:S</th>
<th>I:P</th>
<th>P:S</th>
<th>AARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-2006</td>
<td>1.9:1</td>
<td>2.2:1</td>
<td>1:1.2</td>
<td>Pre-study</td>
</tr>
<tr>
<td>Group A &amp; B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007</td>
<td>1:2.8</td>
<td>1.3:1</td>
<td>1:2.3</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group C &amp; D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td>1:2.8</td>
<td>1:1.7</td>
<td>1:1.4</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group E &amp; F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td>1:2.1</td>
<td>1:1</td>
<td>1:2.1</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group G &amp; H</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010</td>
<td>1:5</td>
<td>1:3.1</td>
<td>1:2.4</td>
<td>Post-study</td>
</tr>
<tr>
<td>Group I &amp; J</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


This table, based on the chi-square test results (see Appendix 18), offers comparison both horizontally, across any one particular year, and vertically, throughout the

---

30 These frequency ratios are derived from chi-square tests for the 1st Writing Assignment which can be viewed in Vol. 2, Appendix 18.
duration of the AARP. Looking across the Pre-study year, 2005-2006, it is clear that in comparison with the Self-assessors and the Peer-assessors, the Instructor-assessor has rated this 1st Writing assignment more leniently. Peer-assessors, on the other hand, have rated similarly to Self-assessors, with a value of 1:1.2. Looking vertically down the Peer-assessor:Self-assessor (P:S) column, there is consistent and close coincidence between these raters in 2005-2006 and 2007-2008, with 1:1.2 and 1:1.4, and there is also consistency, but greater divergence, between them in 2006-2007, 2008-2009 and 2009-2010 with 1:2.3, 1:2.1 and 1:2.4 respectively.

Scanning the I:P column, it can be determined that the Instructor-assessor was rating leniently in comparison with Peer-assessors in the Pre-study, 2005-2006, with a ratio of 2.2:1 and this same pattern is repeated in the following year, 2006-2007, but with a reduced ratio of 1.3:1. In 2007-2008, the I:P ratio is both reversed and increased to 1:1.70, while in the following year, 2008-2009, it equalises at 1:1, only to rise again in 2009-2010, but well beyond the value of the year 2007-2008, to 1:3.1. What is most notable is the complete reversal in rating behaviour in the Post-study, in 2009-2010, as compared with the Pre-study, in 2005-2006, which, due to the Intervention exercises conducted in 2009-2010, one would have expected to have evolved differently.

The equalisation effect of the I:P ratio in 2008-2009, previously mentioned, is clearly mirrored in the equilibrium of the S:I and P:S ratios, with equal values of 1:2.1. This equalisation can be seen very clearly in Figure 5.1 below:
The same pattern of reversal, which was noted in Instructor:Peer (I:P) rating behaviour, occurs between the Instructor-assessor and Self-assessors, with an I:S value of 1.9:1 in the Pre-study year, 2005-2006, revealing comparative Instructor leniency in rating, which is dramatically reversed in the Post-study to a value of 1:5. In the intervening years, the same high ratio of 1:2.8 is repeated in 2006-2007 and 2007-2008, with this dropping to 1:2.1 in 2008-2009, making the sharp rise in the Post-study year, 2009-2010, all the more difficult to justify, but similarly implying negative outcomes from Intervention processes. These reversals in rating behaviour between the Pre-study and the Post-study are colour-highlighted in Table 5.30.

There are three main conclusions which can be drawn from the rating patterns displayed in Figure 5.1. Firstly, in the Pre-Study, S:I ratios and P:I ratios, were very similar. Secondly, these same ratios continued to follow a very similar pattern of rises and falls throughout the AARP, although for the most part the S:I ratios are more exaggerated. Thirdly, it seems clear from the sharp rises in these ratios in the Post-Study, in 2009-2010, that the IE appeared to have a negative, rather than the hoped-for positive, effect on P-A and S-A rating.
5.4.2 Rating frequency ratios for the AARP 2nd writing assignment, 2005-2010

Rating frequency ratios derived from the chi-square test results for the assessment of the 2nd Home Writing Assignment (see Appendix 19) are worthy of scrutiny since they differ quite significantly in a number of ways from those of the 1st Home Writing Assignment. Again, in Table 5.31 below, there is interest in the horizontal perspective across any one particular year of the AARP as well as interest in the vertical perspective throughout the duration of the study. The ratios for the Pre-Study year, 2005-2006, when we compare Table 5.31 with Table 5.30 reveal a complete reversal for I:S, with the ratio 1.9:1 now 1:1.9. The ratio for I:P which was 2.2:1 has now almost equalised, with 1.1:1, while at the same time the previous P:S ratio of 1:1.2 has risen slightly to 1:1.8.

Compared to the Main Study and the Post-Study, the ratios in all categories in the Pre-Study, 2005-2006, show the greatest restraint and overall closeness. Restraint in the 2nd cycle of writing assessment can also be noted in the Post-Study, both in the I:S value which has been reduced from 1:5 to 1:1.7 and also in the I:P value which has reversed since the 1st Writing assessment cycle, from 1:3.1 to 1:5:1. Again, this seems to indicate a delayed reaction to the Intervention exercises in the 1st assessment cycle, with feedback on assessment gleaned from that cycle feeding forward into the 2nd assessment cycle (Murlagh & Baker, 2009), a phenomenon described elsewhere in relation to Oral assessment (Cheng & Warren, 1999).
Table 5.31  Rating frequency ratios for the 2nd Writing Assignment

<table>
<thead>
<tr>
<th>Group &amp; year</th>
<th>I:S</th>
<th>I:P</th>
<th>P:S</th>
<th>AARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-2006</td>
<td>1:1.9</td>
<td>1:1.1</td>
<td>1:1.8</td>
<td>Pre-study</td>
</tr>
<tr>
<td>Group A &amp; B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007</td>
<td>1:6.6</td>
<td>1:1.5</td>
<td>1:3.1</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group C &amp; D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td>1:2.6</td>
<td>1:1.2</td>
<td>1:2.5</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group E &amp; F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td>1:2.8</td>
<td>1:1.0</td>
<td>1:2.4</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group G &amp; H</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010</td>
<td>1:1.7</td>
<td>1:5:1</td>
<td>1:2.6</td>
<td>Post-study</td>
</tr>
<tr>
<td>Group I &amp; J</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


The reasonably consistent I:S ratios from the assessment of the 1st Writing Assignment remain so in the 2nd Writing Assignment assessment in 2007-2008 and 2008-2009, with 1:2.6 and 1:2.8 respectively, but in 2006-2007, for some reason, it has peaked sharply, with a ratio of 1:6.6, whereas the Post-Study ratio of 1:1.7 has fallen and is close to the Pre-Study value of 1:1.9, as illustrated also in Figure 5.2 below:

![Figure 5.2](image)

2nd WRITING ASSIGNMENT

Figure 5.2  y-graph of 2nd Writing Assignment rating behaviour (2005-2010)

31 These frequency ratios are derived from chi-square tests for the 2nd Writing Assignment which can be viewed in Vol. 2, Appendix 19.
Two other changes in the 2006-2007 ratios are also worth mentioning. The I:P ratio of 1.3:1 for the 1st assignment is reversed in the 2nd writing assessment cycle to 1:1.5, and while, as previously mentioned, there is very strong differentiation between I:S ratios, with 1:6.6, there is also differentiation between Peer- and Self-assessors, with a ratio of 1:3.1, which has risen since the 1st cycle from 1:2.3, which can clearly be seen by comparing the y-graphs in Figures 5.1 and 5.2 above.

The I:P column for the 2nd assignment in Table 5.31 shows a pleasing consistency from 2005-2009, with ratios equal (1:1), in 2008-2009, or close to equalisation in 2005-2006, with 1:1.1, and in 2007-2008, with 1:1.2, as well as a ratio of 1:1.5 in 2006-2007 which is reversed in the Post-Study, to 1.5:1, revealing greater leniency in assessment that year on the part of the Instructor and greater severity on the part of Peer-assessors. This closeness and consistency in I:P assessment ratios is also clearly illustrated in Figure 5.2, as are the falls in both I:P and I:S ratios in the Post-Study, 2009-2010, exhibiting, in the latter case, not just a return to Pre-Study ratios, but indications of even greater restraint. The rising peaks for all ratios in the first year of the Main Study, 2006-2007 are also apparent, particularly in the case of I:S, which spoils what would otherwise be a reasonably pleasing, consistent rating pattern throughout the AARP.

As in the statistical analyses provided in the first three sections of this chapter, it is very difficult to reach strong conclusions from the frequency ratios derived from Chi-test analyses that assessment of writing in the 2nd Writing assignment assessment Cycle was significantly better or worse than in the 1st Writing assignment assessment Cycle, but there do seem to be signs of greater regularity, consistency and restraint in rating, overall, with the exception of I:S, 2006-2007, as previously mentioned, in the 2nd Writing assessment Cycle.
5.4.3 Rating frequency ratios for AARP Oral Assignment assessment, 2005-2010

The rating frequency ratios derived from the chi-square test results (see Appendix 20) for the assessment of Oral assignments in the AARP are given below in Table 5.32. Looking across the years on the horizontal level, it can be observed that the year with greatest equilibrium between ratios is that of 2008-2009, with an I:S ratio of 1:1.8, an I:P ratio of 1:1.3 and a P:S ratio of 1:1.5. There is also equilibrium in the Post-Study year, with an identical I:P value to that of 2008-2009, while the ratios for I:S are higher, at 1:2.1 and for P:S are 1:2.3. The I:P ratios just mentioned are equivalent to the 2006-2007 ratio of 1:1.4, while the Pre-Study year and the Main Study year, 2007-2008, produce the same I:P ratio of 1:2.1, as shown in Table 5.32 below:

Table 5.32 Rating frequency ratios for Oral Assignment assessment

<table>
<thead>
<tr>
<th>Groups &amp; year</th>
<th>I:S</th>
<th>I:P</th>
<th>P:S</th>
<th>AARP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-2006</td>
<td>1:1.2</td>
<td>1:2.1</td>
<td>1:3:1</td>
<td>Pre-study</td>
</tr>
<tr>
<td>Group A &amp; B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007</td>
<td>1:2.7</td>
<td>1:1.4</td>
<td>1:1.7</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group C &amp; D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2008</td>
<td>1:4.0</td>
<td>1:2.1</td>
<td>1:1.9</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group E &amp; F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-2009</td>
<td>1:1.8</td>
<td>1:1.3</td>
<td>1:1.5</td>
<td>Main Study</td>
</tr>
<tr>
<td>Group G &amp; H</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-2010</td>
<td>1:2.1</td>
<td>1:1.3</td>
<td>1:2.3</td>
<td>Post-study</td>
</tr>
<tr>
<td>Group I &amp; J</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Regarding I:S ratios, the Pre-study value is pleasingly close to equalisation with a value of 1:1.2, which rises sharply to 1:2.7 in the following year, 2006-2007, and continues to rise steeply, reaching the value of 1:4.0 in 2007-2008. This drops sharply to 1:1.80 in 2008-2009 and rises slightly to 1:2.1 in the Post-Study. These accentuated rises and consequent falls can be seen clearly in the y-graph in Figure 5.3 below:

---

32 These frequency ratios are derived from chi-square tests for the Oral Assignment which can be viewed in Vol. 2, Appendix 20.
Figure 5.3 *y*-graph of Oral Assignment rating behaviour (2005-2010)

The P:S ratio for Oral assessment in the Pre-Study is 1.3:1, while this is reversed in all consequent years, with values of 1:1.7 in 2006-2007, 1:1.9 in 2007-2008, and 1:1.5 in 2008-2009, which rises in the Post-Study to a value of 1:2.3.

Figure 5.3 above, through its peaks for all ratios in the year 2007-2008, offers convincing evidence of the vagaries of ‘rogue’ Group E, while the close agreement between I:S, I:P and P:S, in 2008-2009, shows coincidence which we might have more readily expected to have occurred due to Intervention in 2009-2010. Another interesting pattern is the I:S ratio in the Pre-Study, which, with the exception of 2008-2009, never came close to repetition, but, rather, came to be usurped by the I:P ratio in 2006-2007, 2008-2009 and 2009-2010, and in all cases is close to equalisation, which is pleasing.

5.5 Concluding Remarks

In this chapter, the findings of the AARP have been recorded and presented. It is important to remember that what was outlined and described above showed the processes involved in this teaching and assessment research study, aimed at
promoting learner autonomy through assessment processes, and it was these very processes which were of importance and significance rather than any product or statistical outcome (Chen, 2008; Dickinson, 1987; Luoma & Tarnanen, 2003; Somervell, 1993). Clearly, there was great learning potential in this project and it was up to individual students how much they were willing to commit themselves and exercise their deeper thinking processes which brought self-knowledge, self-direction and greater self-determination. Individual learners differed with respect to the ‘degree’ to which they developed these capacities (Little & Dam, 1998), which are the key to achieving autonomy.

In this chapter, the assessment outcomes and the assessment questionnaire results were presented for the AARP Pre-Study, the Main Study and the Post-Study and where there seemed to be evidence of research questions having been answered, this was indicated. In the next chapter, the findings which have been presented in this chapter will be further analysed and will also be compared with the findings of other researchers in the field. Evidence from quantitative and qualitative data will be used to support assumptions, as will evidence from other studies. This will be done in an attempt to show that learners involved in the AARP were much less Self- and Peer-assessment experimental guinea-pigs and indeed much closer to being members of a community of “active learnership” (Natri, 2007, p. 108), with extraordinary insights and self-awareness, which demonstrated that they were apprentices (Falchikov, 2007, p. 135) in the craft of assessment and of language learning (Nunan, 2000, p. 5) who were master practitioners, or experts, in-the-making.
Chapter 6
Discussion of AARP results

6.0 Introduction

In this chapter, the findings and results presented in Chapter 5 will be discussed and analysed. In the first section, the outcomes of the two cycles of writing assessment which were conducted in the three stages of the AARP will be considered, and particular phenomena which arose from these will be closely examined and interpreted, taking into account findings from other writing assessment studies. This will be followed by discussion of the findings of the oral assessment cycle of the Pre-Study, Main Study and Post-Study stages of the AARP, noting any similarities and differences in behaviour of the cohorts within each stage of the study, as well as across the various stages, and in some cases oral assessment behaviour will be compared with writing assessment behaviour. Similar studies of oral assessment from the international literature will be referred to in order to compare and contrast their findings with those from the AARP. Finally, it will then be determined how successful the AARP was in answering the four research questions, which were presented in Chapter 1, Section 1.4.

In formulating this discussion, it is important to remember that the bulk of the evidence recorded in this thesis can only be presented in the form of the recordable ‘products’, which are the scores produced in the assessment process and the ratios produced by answers to questionnaires. This gives us access to the ‘mechanics’ of scoring and rating and the ‘bare bones’ of the project which can be viewed in the statistical data, whereas what was most important was actually the ‘processes’ which were stimulated by the AARP, which in some students resulted in a transformation, which even they were not aware of. The fact that some students may have rejected this approach to assessment, does not mean that it was a failure, as
these students were still challenged to question and think, much more than if their grades had simply been provided by the instructor in the traditional way.

It is hoped that at least some of the evidence will show that the students involved in the AARP were more than just passive recipients of knowledge or “pawns” (deCharms 1997, cited in Stefanou et al., 2004, p. 98), but, for the most part, were engaged in transformative processes enabling them to become “origins” (ibid), “agents” and “experts-in-the-making” (see Chapter 2, Figure 2.9). The chapter will end with the conclusions which can be derived from the AARP overall.

6.1 Writing assessment in the AARP: Discussion and analysis
In the sections that follow, the outcomes of the two writing assignments will be discussed and then a number of interesting cases and phenomena which emerge from these two cycles of writing assignment assessment will be outlined. To complete the discussion of writing assessment, comparisons with other writing assessment research studies will be made.

6.1.1 Outcomes of Assessment of the LM I 1st Home Writing Assignment
Our initial impressions from a reading of the writing assessment outcomes from the three Studies (Pre-, Main and Post-) within the AARP, based on comparison of the true-score means between Self-assessment, Peer-assessment and Instructor-assessment, for the 1st Home Writing Assignment (see Vol. 2, Appendix 17) are ones of disappointment. While the two research groups in the Pre-Study (2005-2006) seemingly performed well, with no significant differences in assessment in either case, in the three consecutive years of the Main Study, we see that the first of each of the two groups in each year has produced significant differences, and this same pattern is repeated in the Post-Study (2009-2010), both when we take all the
participants’ assessments into account in that semester and also when we exclude those who did not take part in the Intervention exercises.

If, in estimating assessment success, we judge purely from the groups in which there are no significant differences between Self-, Peer- and Instructor-assessment, then taken overall it can be claimed, with certain reservations, that Groups A, B, D, F, H and J appear to have enjoyed success in their assessment, giving an overall success rate of around 60% as compared with the total student constellation over the five-year period.

What is interesting, from a comparison of the mean scores over the five-year period is that in 7 out of 10 cases, the P-A mean score is lower than the I-A mean score, showing greater severity on the part of Peer-assessors as compared with the Instructor-assessor. In 8 out of 10 cases, the S-A mean score is the highest, revealing leniency in rating and therefore a lack of modesty, a phenomenon which conflicts strongly with other published studies of writing self-assessment (Matsuno, 2007; 2009; Saito & Fujita, 2004). In the other two cases, in Groups A and B, in the Pre-Study, the I-A mean score is the highest, showing greater leniency in the rating of the instructor, as compared with P-A and S-A, a phenomenon which never re-occurred for the duration of the AARP. In no cases is the P-A mean score the highest, thus revealing overall strong alignment, on the part of Peers, with the Instructor’s rating.

With regard to Writing assessment success, if we also take into account the Pearson Product Moment correlation coefficients (see Appendix 17), the picture looks a little more favourable, particularly in the case of Group C, 2006-2007, where we have correlation coefficients both in the case of Self-Instructor (S-I), and Peer-Instructor (P-I) (see Table 5.8 and/or Appendix 17), which are very encouraging and help to cancel out our previous concerns to some extent. The S-I
correlation coefficient for Group F is also moderate confirmation that things went well, while the P-I correlations for Groups I and J in the Post-Study (see Table 5.22 and/or Appendix 17) with Intervention are also heartening, particularly in the case of Group I, though, strangely enough, the former correlation is lost if our calculations are based only on the 18 participants who were trained in peer-assessment through Intervention, while in the case of Group J, if we restrict our statistical analysis to the 18 trained participants, then the P-I correlation coefficient rises slightly (see Table 5.22).

In judging these results of assessment of the 1st Home Writing Assignment, it should be remembered that this was a first attempt at writing assessment for all participants and, that, with the exception of the students involved in the Post-Study, no training had been offered or given. The pedagogical philosophy behind this approach was that through the experience and processes of Peer-assessment, albeit one-off, using it as a form of what Falchikov refers to as “apprenticeship” (2007, p. 129), learners would be sufficiently practised to enable them to conduct Self-assessment. Indeed, based on his experience and involvement with learner training both for writing and oral skills assessment, Saito (2008, p. 576) is led to conclude that “longer training alone may not provide further improvement in correlation” and that Peer-assessment is “fairly robust”, by which he means that Peer-assessment seems to be sufficiently reliable in itself without very much training preceding it, if at all.

While the results, for a first attempt at writing assessment (see Vol. 2, Appendix 17), in the AARP, look favourable compared with studies of Writing assessment by Matsuno (2007; 2009) and Saito & Fujita (2004); nevertheless, the results for Groups E and G, in particular, remain disappointing and puzzling, and perhaps warrant further investigation as to possible parameters which created more
outliers or stronger outliers in these two groups, which led to these deviations. Rühlemann (2002, p. 4) takes the interesting view that it is “futile” for teachers to expect learners to evaluate writing samples in the same way as themselves and that “co-evaluation entails compromise”. Such co-evaluation is invaluable since it “establishes writing as a communicative act” (2002, p. 8), while Hu (2005, p. 324), in speaking of peer review, believes that it promotes “a sense of text ownership”, both of which were potential benefits in the AARP.

Possible parameters involved in the 1st Cycle of writing assessment will be discussed after comparison with the outcomes from the 2nd Cycle of writing assessment, which follows.

6.1.2 Outcomes of Assessment of the 2nd Home Writing Assignment
In examining the results of the 2nd Writing Assignment assessment, it will be interesting to see if suggestions made by researchers such as Lim (2007) and Nakamura (2002), who imply that the more that assessment is carried out by learners, the more accurate their assessment will become, is borne out by evidence gathered from the 2nd Cycle of writing assessment in the AARP. If, indeed, assessment improves with practice, which, in this case involved a second attempt at Peer-assessment and a second attempt at Self-assessment, then we would expect to find signs of amelioration in these assessment processes, particularly among the seemingly less successful groups in the 1st Cycle. However, it has to be remembered that both the production and the assessment of the 2nd Home Assignment was necessarily more demanding and required greater effort on the part of learners, since new assessment criteria also had to be assimilated (see Appendix 8).

An initial appraisal of the outcome of assessment processes based on comparing the means of true scores from S-A, P-A and I-A of the 2nd Writing
Assignment, at first sight appears to be more disappointing than the 1st Writing Assignment. While the same behavioural patterns occur in both assessment cycles in the Pre-Study, in 2005-2006, with no significant differences in either Group A or Group B, indicating strong learner and instructor assessment alignment, at the same time, assessment behavioural patterns are repeated in both Writing cycles in the Post-Study (2009-2010), with significant differences in Group I and no significant differences in Group J, indicating less assessment alignment in Group I. From thereon, the similarity in assessment patterns ends and instead we have significant differences in five out of the six groups in the Main Study, with the same pattern of A>B=C occurring in three of those, namely Groups C, D and E, indicating that over-inflated self-assessment has been to blame for non-alignment, with closer alignment between Peer and the I-R. Our first impression therefore is of something approximating to 40% success overall in this second attempt at formative assessment of writing.

Interestingly, a comparison of the mean scores over the five-year period produces a different picture from that for the 1st Writing Assignment, since, in the case of the 2nd Writing Assignment assessment, in 10 out of 10 cases, S-A repeatedly produces the highest mean score, displaying consistent leniency and lack of modesty in Self-rating, which again conflicts with the results of other previously-mentioned published studies, where S-A deviations are due to over-modesty, so that in no cases in the AARP do P-A or I-A produce the highest mean score. In 3 out of 10 cases, the mean score for P-A is less than I-A and in 1 out of 10 cases, the mean scores for P-A and I-A are the same, revealing a tendency towards more severity or similar severity in assessment on the part of the Peers to that shown by the Instructor, which is interesting.
As with the 1st Writing Assignment, the assessment picture for the 2nd Writing Assignment changes when we take correlation coefficients into consideration. S-I correlations occur in three of the groups from the Main Study which had previously produced significant differences, namely Groups C, G and H (see Appendix 17), which is a very encouraging sign. In addition, in the Post-Study, Group J has a correlation coefficient occurring between Peer and Instructor (see Appendix 17), which is also satisfying. Depending on how optimistically one regards these correlations, it could be claimed that this brings the level of assessment success to the same level as in the 1st Writing Assignment, which is approximately 75% over the five-year period.

With regard to correlation coefficients, Saito and Fujita (2009) make the rather tongue-in-cheek remark that the “beauty” concerning the “degree of strength” in these tends to lie “in the eye of the beholder” and, likewise, Porte (2002) warns us that with small groups of subjects, individuals yield more strength as outliers and can therefore cause more distortion. Bachmann (2004, p. 253) also reminds us that their “meaningfulness” relates not only to the “size”, but also the “hypothesised relationship” between the two variables, while Schelfout et al. (2004) assert that “correlations can only give a relative view of similarity and not an absolute one”. Perhaps it is a case of regarding the proverbial ‘glass’ of assessment as half-full or half-empty in deciding if these results are the same or different from the 1st cycle. Besides, Dörnyei (2007, p. 223) reminds us that in the field of Linguistics, even low correlation coefficients can be very meaningful.

Another interesting feature of the 2nd Writing assessment cycle, which deserves mention, is the fairly consistent rating frequency ratios for I:P, derived from chi-square tests (see Appendix 19), throughout the AARP (see Chapter 5, Table 5.31 and Figure 5.2), revealing what seems to be much greater alignment and
consistency between Peer-assessment and Instructor-assessment than in the 1st Writing Cycle. Such rating consistency is surprising in a task which was necessarily more demanding, as was the assessment of it and could be an indication of progression in “learner assessment autonomy” (Falchikov, 2007, p. 140).

Unlike the AARP, although the research studies of Matsuno (2007; 2009) and Saito & Fujita (2004) involved two cycles of writing assessment, neither study offers two sets of results from the same groups for comparison, which would have been interesting.

Even if it can be claimed that the results of the two cycles in the AARP are similar, and that the results show promise, there is perhaps justification in being concerned about the changes in assessment behaviour, compared to the 1st cycle of assessment, so that it is possible that, in some cases, learners were introducing their own criteria and rating-scales and rewarding themselves for effort (Fallows & Chandrahamon, 2001, p. 234). There seem to be no saving graces in the case of Groups D and E and, particularly in the case of Group D, the changes in assessment behaviour since the 1st cycle of Writing assessment, seem to warrant explanation. In fact, the instructor did make a slight change in assessment procedures with Groups C and D in 2006-2007 with the 2nd Writing assignment, which may either help to explain the phenomenon or add to its seeming complexity.

6.1.3 Special cases and phenomena in AARP writing assessment
In this section, a number of special cases and phenomena which emerge from the two cycles of Writing assignment assessment and which seem worthy of further investigation will be brought to the surface and examined under the microscope. The phenomenon with which this investigation will commence is the one just
previously referred to, concerning a change in Writing assessment procedures in the 2nd Cycle, and the resulting outcome, in 2006-2007.

6.1.3.1 Self-assessment maturity

In Chapter 4, Section 4.6.2.1, a variation in the procedures for Self-assessment was outlined where students were asked to study the assessment criteria for the 2nd Home Writing Assignment carefully before submission of the assignment and, breaking with precedence, to submit a completed Self-assessment form together with the submission of their assignments. This occurred in the 1st year of the Main Study, 2006-2007, when 20 students from Group C and 13 from Group D conformed to this request. These first Self-assessment ratings were recorded, and from thereon assessment procedures were carried out in the normal way. After the passing of an appropriate period of time, the usual procedure with Peer-assessment was followed, in which 22 students from Group C and 18 from Group D participated, and some time after that the students were requested to assess their own assignment (Self-assessment), which for some constituted a second Self-assessment rating. The outcome of this break with assessment tradition can be seen in Chapter 5, Section 5.2.3.1, and, specifically, Table 5.13.

What is quite bewildering is the fact that so many students should have awarded themselves a higher grade in their second Self-assessment, after having been through the processes of Peer-assessment, and it is very hard to find a satisfactory explanation for this behaviour. One explanation that could be offered is that there was sufficient assessment ‘maturity’ (Ritter, 1998, p. 79) before Peer-assessment processes for these students to Self-assess themselves reasonably accurately, indicating that practice through Peer-assessment with these groups led to
less rather than more Self-assessment accuracy, leading them to deviate further from Instructor-assessment.

The majority of experts regard Peer-assessment as something of a springboard, with it being described as a “stimulus” (Black & Wiliam, 2006, p. 13) or as a “prelude” (Alderson, 1994, p. 10) to Self-assessment; in other words, they see the necessary sequence as being to perform Peer-assessment first, where there is less emotional overload or “interference” (Taras, 2001, p. 609) involved, and for this to be followed up by Self-assessment. Interestingly, Leblanc & Painchaud (1985, p. 675) see Self-assessment as a “fine-tuned” facility which they say does not come “automatically to all students”, but thereby implying that it might come naturally to some. Also, unlike most assessment researchers, Pope (2005, p. 52) takes the rather unique stance that the sequence of assessment should be the other way round, with Self-assessment experience being the necessary prerequisite for the conducting of Peer-assessment, so that “the two activities inform each other”. Perhaps this implies that as instructors we should be aware that the ability to Self-assess may, in some cases, be an innate ability, or that some individual students or groups of students may have somehow already matured sufficiently to enable them to conduct Self-assessment with accuracy, without the intermediary step of Peer-assessment. At the same time, instructors should perhaps also be receptive to the two approaches to learner-assessment, with Peer-assessment preceding Self-assessment or vice versa, and try to find ways of determining which sequence is more appropriate for a particular individual or group of individuals. In the literature, there are no examples of studies where Self-assessment preceded Peer-assessment, but clearly this warrants further investigation. It is otherwise hard to explain why the students in Groups C and D of the AARP overrated themselves to the extent that they did after Peer-assessment of the 2nd Writing assignment (see Chapter 5, Table 5.12).
In the next section, the phenomenon which will be discussed is that of the Writing assessment performance of Groups A and B in the Pre-Study, which was never repeated.

6.1.3.2 The unique success of the Pre-Study
Disappointingly, the dual success of Groups A and B in both cycles of writing assessment, with the means of both groups in both assignments having no significant differences, showing alignment between self-, peer- and instructor-assessment, was never repeated in the remaining four years of the AARP, by any other pair of groups, although it was repeated individually by Group F in the Main Study and Group J in the Post-Study (see Appendix 17). This raises questions of what might have been unique to the two groups in the Pre-Study, in terms of some kind of positive variable which was never again reproduced or replicated for the duration of the AARP.

In the Pre-Study it should perhaps be stressed that learners were trusted to take all the paperwork related to writing assessment home and conducted both Peer- and Self-assessment, on separate occasions, in the privacy of their home environment and it may have been that this helped them to feel that they were taking on the role of the Instructor, enabling them to develop better what Nunan (1988, p. 134) refers to as “a critical self-consciousness”, which made it possible for them to adopt the role of “active agents within the learning process” and, in the case of Self-assessment, were quickly able to apply what Heron (1981, p. 64) deems “the art of balance between self-denigration and self-inflation” and what Stickler et al. (1999: 285) see as “a balance between criticism and tolerance”. In cases where such a “balance” has not been achieved, Dickinson (1987, p. 137) reminds us that we can expect “different degrees of involvement of ‘the self’”. Some researchers, like Chen

33 What was also remarkable was that Peer-assessors rated more strictly than the Instructor.
(2006a) and AlFallay (2004), by way of explanation for such phenomena have suggested psychological and affective influences on learner assessment behaviour, factors which both Kohonen (1999) and Taras (2001) also consider important.

When the results of the Pre-Study are compared with the Main Study, then it becomes clear, as Freeman (1995, p. 298) states, that when one cohort of students produces reliable results, it does not mean to say that “subsequent cohorts of students will be equally reliable”. Fallows and Chandrahamon insist (2001, p. 234) that in order for such a system of formative assessment to be successful that there must necessarily be “a significant level of mutual trust”, not only between the students and the instructor, but also “between the students themselves”. It is possible that this level of trust, in the case of the Pre-Study, was increased by permitting learners to take the paperwork home, something which in years previous to the AARP had proved disastrous, since many assignments were lost by students or never returned.

While the overall element of success in the AARP, from the aspect of Writing assessment seems to be evidence of such trust, it is possible that the change to in-class Peer- and Self-assessment of Writing, in the years following the Pre-Study, as opposed to assessment at home, reduced students’ sense of focus and privacy, leading to undercurrents of pressure from peers, who were sitting next to them, and possibly also what Ritter (1998, p. 79) refers to as the “corrupting effects of personal power”.

Another interesting point to note is one way in which the two groups, A and B, in the Pre-study of 2005-2006 differed from the pairs of groups in the years that followed, and that is that they were distanced alphabetically, with another group of Language Mastery I intervening, so the students in these two groups, unlike the others, did not have contact with the other participants on their remaining
compulsory 1st year courses. In the years 2006-2010, on the other hand, all of the
Groups C, E, G and Group I had contact with a large number of the students in
Groups D, F, H and Group J, correspondingly, for Introductory Literature and
Linguistics courses and for that reason there is no doubt that they consequently had
formed friendships and a sense of camaraderie between them. In the Pre-Study there
was no such intermingling, which meant that their peers in the second group were
less familiar to them and so they could perhaps view the assignments of those peers
with greater impartiality and treat the assessment process with much greater
objectivity and equilibrium in a less emotionally-charged environment, emotion
being a factor, already mentioned, which can interfere with assessment processes
(Jones & Wiliam, 2008, p. 21; Taras, 2001, p. 609). The point being made here is
that the size of student groupings for other subjects may have caused friendships to
be formed between members of the two LM I groups used each year in the AARP
Main Study and the Post-Study, which could have resulted in some cases of
“friendship marking”, particularly amongst members of the first, and in each
instance, larger of the two groups. The matter of group size is mentioned somewhat
in passing by Falchikov & Goldfinch (2000, p. 315), who suggest, based on their
review of the literature, that comparison of the ratings of groups of more than 20
individuals, may lead to “poorer agreement”. The matter of the number of students
in LM I classes is a concern which will be discussed in the next section.

6.1.3.3 Variables within large groups
An interesting point to note is the size of the research groups which, with the
exception of the two groups in the Pre-Study, may also have had an effect on
learners’ Writing assessment behaviour. From the data provided in Chapter 4, Table
4.4, it is apparent that the first of the two groups in each year of the AARP is always
larger than the second. While in the case of the Pre-study, this was purely a chance
occurrence which seems to have had no particular effect, in all other years, the
larger number was attributable to the fact that the first and second alphabetically-
sequenced groups in each autumn semester were those assigned to the I-R to teach,
and it was customary for administrative staff to assign a large number of
‘problematic’ cases to the first alphabetical group in each year. By ‘problematic’ it
is meant that students were re-registered on the course for various reasons and
therefore tended to be older. This difference in age is particularly noticeable in
Groups E and G (see Table 4.4), which each have the highest number of students of
any of the participating groups in the AARP, with 30 students in each case, and
while each has a fairly average number of students aged 17-19, at the same time
each has a particularly high ratio of students aged 20 and over. These ratios stand at
20:10 and 18:12 in the case of Groups E and G respectively.

Interestingly, Harris (1997, p. 18) makes the remark that in the case of Self-
assessment that “younger learners” may very well be “less resistant” to this kind of
cooperative assessment than their older counterparts, so it is possible that age and
“conditioning” were the significant factors in the present case, with learners
suffering from what Boud (1996, p. 44) refers to as “debilitating dependencies on
others” and what Seligman (1975, cited in Williams & Burden, 1997, p. 130) and
Dörnyei (1994, p. 276) call “learned helplessness”. There is also the matter of
“response effects”, to which Heilenman (1990, p. 175) draws our attention, where
students “may distort or suppress information” for reasons of “acceptability and
self-presentation”, or of “halo effects”, described in Feeley (2002) as well as Nisbett
& Wilson (1977). It seems likely that the older students came to these learning and
assessment processes with much stronger expectations and beliefs, much deeper-
rooted “cultural” convictions (Harris, 1997, p. 12; Valdez Pierce, 1999, p. 131) and
preconceptions which it was not possible to change over the course of 10-13 weeks. Taras (2001, p. 608) regards time as all-important if students are to master the “protocols and procedures” of assessment, while Allwright (1990, p. 7) considers that autonomy will also “take a long time to develop”.

Chen (2006a, p. 14), in a similar study of self- and peer-assessment to the AARP, but with oral skills, amongst 2nd year English majors at a University in S. Taiwan, encounters such attitudes which she believes “years of schooling might have conditioned” which has made learners reluctant to see the teacher in anything but “an authoritative role”. Black et al. (2003, p. 98) also warn that even high-attaining students may be resistant to innovation as they become used to simply “getting the answers right” and they do not welcome changes; likewise, those who are used to simply being “tacit observers” may resent being asked to make more effort. Black et al. (ibid) even cite the case of a teacher who with all his classes had encountered no difficulty in introducing formative assessment until he met one class where they were so accustomed to passivity that he found “it was impossible to change their attitude to learning”.

An educational system and curriculum which promotes heteronomy (Waite-Stupiansky, 1997, p. 23; Peters, 1998, p. 85), and thus “learned helplessness” (Dörnyei, 1994, p. 276; Seligman, 1975, cited in Williams & Burden, 1997, p. 130), rather than autonomy, causes learners to become accustomed to “spoon-feeding”, even in a university environment, prompting McKay & Kember (1997, p. 55) to complain of students in higher education in Hong Kong that “[t]hey wish to be spoon-fed and in turn they are spoon-fed”. It is for this reason that there have been moves in Hong Kong, the U.K. and elsewhere (Beaman, 1998; Biggs, 1999; Clarke, 1998; Dochy, 2001; Dochy, Segers, Gibjels & Struyven, 2007; Kember & Gow, 1989; Knight, 2006; Murphy, 2008; 2011; Sluijsmans et al., 1999; Zariski, 1996)
towards a system of assessment which encourages greater learner involvement and ownership. Students who are products of this very common type of educational system may put up resistance to changes which instructors try to introduce, particularly where the matter of assessment and grading is concerned. Knowles (1975, p. 85) points out to us that traditional educational methods foster a sense of competition for grades, which he believes automatically puts adults entering education into a “defensive frame of mind”. They consequently feel threatened by a “self-diagnostic process” designed to identify “one’s weaknesses” and indicate the need for “additional learning”. It is a process they find “both strange and threatening”. It is such phenomena which will be discussed in the next section.

6.1.3.4 Rogue groups

It was noted in the previous section that Groups E and G were the two largest cohorts of the total ten in the AARP and that age may have been a factor in their assessment behaviour and performance. It will become apparent after discussions of the Oral Assessment outcomes in the AARP that Group E also performed in a similar manner in their Peer- and Self-assessment of Oral skills, making them, effectively, a truly ‘rogue’ group.

A possible explanation is that Group E, in particular, was more willing to give up its autonomy, as referred to by Breen & Mann (1997, p. 141) and preferred to put on a “mask of autonomous behaviour” (ibid), just to pay lip-service to the I-R and comply with her wishes, and still took a “surface” rather than a “deep” approach to their studies. According to Stefani (2004, p. 62), it is not only students who are to be blamed for taking such an approach, as, indeed, instructors do sometimes actually “reward a surface approach to learning”.

247
It is also a possibility that these students viewed their involvement in assessment as what Reynolds and Trehan (2000, cited in Tan, 2004, p. 659), term “a form of governmentality” which was being imposed on them, turning them, in effect into “policemen in self-assessment”. Tan (2004; 2007) warns us of the possible paradox contained within self-assessment, in that though we may think that as instructors we are offering learners more autonomy through involvement in assessment, there is a sense in which they are being subjected to “greater control and surveillance” (Tan, 2004, p. 659). There is, of course, the other equally unpleasant possibility, mentioned by Ritter (1998, p. 79), that these students were prey, as previously mentioned, to “corrupting feelings of personal power”.

Another possible explanation could have been that having to deal with ‘real’ grades in the assessment process caused what Taras (2001, p. 610) refers to as an “emotional reaction” which she believes “interferes with students’ judgements and prevents them from focusing on their work”. This, she feels, also causes students to block out feedback whether from their peers or from their instructor.

This view of dealing with ‘real’ grades comes into conflict with opinions expressed by other researchers such as MacAlpine (1999, p. 16) and Stickler et al. (1998, p. 286) who believe that involving students in the grading process and using peer-derived marks helps to ensure that the assessment process is “taken seriously”. Orsmond et al. (1997, p. 358) assert that:

it is far better to take the risk over the marks than deprive students of the opportunity of developing the important skills of making objective judgments about the quality of their own work (and that of their peers) and of generally enhancing their learning skills.

Little (1999b, p. 6) also believes that learners will take the “business” of assessment more “seriously” if they are involved in grading and contribute to the “final grades”.

248
Harris (1997, p. 18) insists, as Janssen-van Dieten (1989, p. 44) does, that when there are poor assessment results, as there are here in the case of Groups E and G, then this is all the more reason for persevering and involving learners in the assessment process and not abandoning it. On the other hand, Jing (2006, p. 100) is of the opinion that it is “disparity” between the agendas of the instructor and the learners that may result in “learner resistance” and Chen (2006a, p. 14) also seems to believe that in some cases learners simply do not wish to be involved in assessment processes and that it “may not always be what ordinary students desire”.

Of course, students do not always have sufficient knowledge and background to be able to make informed judgements about their abilities and capabilities, as in the case of Miller & Ng (1996), which was discussed in Chapter 3, Section 3.4.2, where students after very active involvement in Peer-assessment took a rather condemnatory stance towards Peer-assessment of oral assignments, deeming themselves unqualified to perform it, since the researchers had not informed them before completion of the questionnaire of their remarkable accuracy, compared to the instructor, in their Peer-assessments. Had they been informed of the outcome of their efforts, it seems highly likely that their views on Peer-assessment would have been modified, although Clifford (1999, p. 122) discovered that this is not necessarily the case.

It is indeed a strange irony that learners who prove themselves to be so capable of assessment should, at the same time, express so much hostility towards it. Faced with “resistance” (Sambell et al., 2006, p. 165) or hostility towards “assessment autonomy” (Falchikov, 2007, p. 139), instructors may decide that Intervention is necessary and it is the case of Intervention in the Post-Study which will be discussed in the next section.
6.1.3.5 The consequences of writing assessment intervention in the Post-Study

The statistical data related to writing assessment in the Post-Study is set out in Chapter 5, Sections 5.3.2 and 5.3.3 and the data includes information concerning the results of the Intervention exercises (IE)s. There is reasonably clear evidence that the IEs themselves worked successfully, although not all students in the two Groups I and J participated in them. Non-participation indicates their absence from class or late arrival for lessons, in a year when attendance became compulsory, and this may be an indication of their attitude to learning in general.

Although both groups performed rather well in the IEs themselves, with Group I performing even better than Group J, when it comes to doing actual assessment as opposed to mock assessment, Group J appears to out-perform Group I in both assessment cycles, and this applies whether we analyse the results only with the participants who underwent training, or not. One saving grace is the fact that both Group I and Group J produce correlation coefficients between P-A and I-A in the 1st Writing Assignment assessment, though that of Group I disappears while that of Group J increases if we include only trained participants (see Appendix 17). When it comes to the 2nd Writing Assignment assessment, Group I once again produces a correlation between P-A and I-A, which rises if we restrict the cohort to trained participants only. Group J, although it produces p values which are not statistically significant, produces no correlations in the 2nd Cycle of assessment (see Tables 5.21 and 5.22). There are therefore some indications that Intervention may, after all, have worked for Group I, but with a delayed effect, and may even have been redundant in the case of Group J.

Taras (2001, p. 608), as already mentioned, makes the very valid point that it is unreasonable to expect learners “to learn marking protocols and procedures overnight” and that it takes time to become part of an “assessment culture” and to
play the role of “double markers of their own work”. It could also be the case that Group I had ‘mixed feelings’ (Saito & Fujita, 2004, p. 36) towards the assessment tasks they had to carry out, or, perhaps, feelings or dispositions that fluctuated.

Findings from other writing assessment research studies
In Chapter 3, Section 3.4, a review of the literature relevant to the AARP was conducted and those which were relevant to self- and peer-assessment of writing were highlighted. Reference was made to the way they conducted their research, the ways in which their data was analysed and the conclusions that were reached. It is hard to make comparisons since the approaches taken by the researchers were quite different and they were concerned with discursive writing rather than descriptive/narrative, and training for writing assessment was involved in all cases. What is noticeable is that although the outcomes of peer-assessment and self-assessment of writing do not always achieve what researchers would expect or hope for, this does not deter them from publishing their findings and in persevering with learner-centred assessment processes, revealing a strong belief on their part that the benefits to be derived from the ‘processes’ of assessment of writing far outweigh the disappointing assessment ‘products’, and are therefore worth all the time and effort invested.

We will now turn our attention to Oral assessment in the AARP.

6.2 Oral assessment in the AARP: Discussion and analysis
The outcomes of the Oral assessment procedures which were followed in the AARP will be discussed in this section and comparisons will be drawn with other researchers’ findings where appropriate.

In our consideration of the degree of success of the Oral assessment Cycle of the AARP, it should be noted that the procedures for Peer-assessment of Oral
presentations were altogether different from Peer-assessment of Writing assignments in that Oral assessment was conducted by multiple peers rather than single peers as occurred with writing (see Chapter 4, Figure 4.1). Self-assessment of oral skills was also different from Self-assessment of writing in that it was conducted immediately after the presentation in class, so that there was no time-lag to permit the “distancing” which Boud (1995, p. 202) thinks is important with assessment of writing. This type of distancing, though useful for both Peer-assessment and Self-assessment of Oral presentations, could only have been achieved if it had been possible to make video recordings of the presentations, as did Chen (2006a; 2006b; 2008), which, in the research setting of the AARP, was not an option. Had such an option been available, it might have been possible to overcome affective factors which could have interfered with and clouded judgment, so that self-rating was based on their “communicative intentions”, rather than what they actually communicated, or the awarding of inflated grades due to the effort expended “to convey messages” (Ross, 1998, p. 9), rather than due to the successful conveyance of those messages. When performing in public in a foreign language, the consequent feelings of excitement, nervousness, relief, satisfaction, or disappointment might obscure objectivity in assessment both in the case of Self-assessment, but also Peer-assessment, particularly when judging co-presenters, an important matter which will be discussed in detail. Interestingly, Ross (1998, p. 8-9) believes that “learners are actually less adept at estimating their own speaking skills”, as compared with other skills, and that their Self-assessment, as mentioned previously, is “quite susceptible to extraneous factors”.

In the AARP, the true scores for Peer-assessment for which the means are calculated and compared with the means of Self-assessment and Instructor-assessment, it should be remembered, are, in the case of Oral assessment, the means
of the means, which is to say that the raw mean peer score which is combined with
the Self-assessment score and the Instructor-assessment score to calculate the
average of the three scores, is already the mean or average score based on the
ratings of all the members of the audience who were present and assessed.

One important way in which the AARP differs from other research studies of
formative Oral assessment is the fact that all class members, rather than either a
selected few or the class subdivided into groups, were permitted to take part in the
Peer-assessment process, albeit a lengthy and time-consuming procedure for the
instructor to tabulate and calculate (Cheng & Warren, 1999, p. 313). Another
important fact is also that there was no discussion in groups after each presentation,
nor was there any immediate feedback given by peers or the instructor to the
presenters. Sometimes the instructor or other members of the audience would ask
questions related to the subject-matter, but on all occasions presenters were thanked
for their contribution and were given an appreciative round of applause, without
comment.

Initial impressions of the AARP Oral assessment results, based on analysis of
the means of true scores from Self-assessment, Peer-assessment and Instructor-
assessment, for the Pre-Study (2005-2006), are the same as with writing, with no
significant differences in either group, and, likewise, there are no significant
differences in any of five out of the six groups in the Main Study, with the one
exception being Group E, 2007-2008, which, it was mentioned earlier, also
performed poorly on assessment of both the writing assignments. Likewise, in the
Post-study, one group, Group I (whole group) appears with significant differences,
while Group J has no significant differences, a pattern which was repeated in both
Writing Assessment Cycles also.
Overall, assessment success with oral skills seems to amount to 80%, or possibly more, as will be explained in the section dealing with the special case of the Post-study, in due course.

Interesting points revealed by statistics

Some interesting statistics which emerge from a survey of the mean scores for the Oral Presentation over the five-year AARP period (see Chapter 5, Tables 5.4, 5.14, 5.26 & 5.27 and/or Appendix 17) are the fact that in 7 out of 10 cases, the mean score of S-A is the highest, with P-A the highest in the remaining 3 cases. This shows that I-A did not have the highest mean score in any case, revealing relative severity in marking throughout the AARP. In 1 case out of 10, the P-A mean score is lower than the I-A mean score, showing greater assessment severity on the part of Peers, and in 1 case out of 10, the P-A and I-A mean scores are almost equal, showing equal severity. Viewed overall, this shows overall leniency in S-A, which is largely shadowed by P-A, with I-A being the strictest in most cases. As with writing assessment, this tendency of self-raters in the AARP to be lenient comes in stark contrast with most studies of oral assessment, which were conducted in the Far East, where self-assessors tend to be overly strict with themselves, for reasons of modesty (Chen, 2006a; 2006b; 2008; Patri, 2002; Saito, 2008).

By taking into account the Pearson Product Moment correlation coefficients (see Appendix 17), on the other hand, there is nice confirmation of assessment success through P-I correlations for 5 of the 10 groups as well as S-I correlations for 3 of the 10 groups. Group F, 2007-2008 (see Appendix 17), has very similar correlations for P-I, S-I and P-S, which therefore comes across as the most successful group of all in assessment, while Group H, 2008-2009, also has both P-I
and S-I correlations which confirm very healthy assessment behaviour amongst its members.

In contrast with the picture derived from the outcome of AARP Writing assessment, in the case of AARP Oral assessment, it is very apparent that the double success of the two Groups, A and B, in the Pre-Study was repeated twice over in the Main Study, with Groups C and D in 2006-2007 and Groups G and H in 2008-2009. This shows consistency in oral assessment behaviour between the groups and this is particularly clear in Groups C and D where there are almost identical correlation coefficients between Peer-assessors and Self-assessors (see Appendix 17) respectively. There is also similarity between Groups C and H in that each produces an identical correlation between Self and Instructor. Group H does even better by producing a higher correlation coefficient between Peer and Instructor (see Appendix 17). The particular cases which do not conform to the standard assessment pattern in Oral assessment in the AARP will now be discussed below.

6.2.1 Special cases and phenomena in AARP oral assessment
The previous section on oral assessment in the AARP highlighted the differences in Peer-assessment procedures with regard to collecting and calculating grades. In the case of Writing assignment assessment, three individuals contributed to the grade awarded for the assignment: the writer, an anonymous peer and the class instructor. In the case of oral assignment assessment, two individuals, in addition to the entire audience, contributed to the grade awarded for the assignment, since assessment was conducted by the presenter, all the peers present and the class instructor (I-R). This same form of simultaneous assessment by the instructor and peers is reported in Chen (2006a; 2006b) and Peng (2010).
The type of peer assessment just described as being conducted in LM I presentations is generally recognised in the literature as being reliable since possible outliers both at the upper end and the lower end tend to be neutralised by the majority, leading to a fair peer grade (Falchikov, 2005, p. 201); nevertheless, there is evidence that in two out of the ten AARP cohorts, this same reliability was not achieved. These two special cases will be examined below, starting with the case of the Post-Study.

6.2.1.1 The consequences of oral assessment intervention in the Post-Study
The aim of the Intervention in the Oral assignment assessment was to offer training and practice in peer-assessment, with a view to possible improvement in assessment performance by the provision of support and scaffolding, consistent with Vygotskian socio-constructivist theory, through mock assessments. The procedures carried out in the Intervention can be found in Chapter 4, Section 4.6.3.1 and the results of the Intervention procedures in the Post-study can be found in Chapter 5, Section 5.3.4.

20 learners from Group I and 18 learners from Group J participated in the oral assessment Intervention exercises, with a view to training learners to be more objective and more efficient assessors of oral presentations, and this, in effect, is what occurred. If we reduce the number of assessors in Group I, from the total 27 who actually conducted peer- and self-assessment to the 20 who undertook Intervention training, then the $p$ value changes to being not significant and, in Group J, by reducing the number of assessors from the total 22 who actually conducted peer- and self-assessment to the 18 who were trained in oral assessment, the $p$ value is increased, but is still considered not to be significant. When only the students who took part in the Oral Intervention Exercise are taken into account, then a Pearson
Quick-moment correlation test also produces a correlation coefficient between Peer-Instructor (P-I) in Group I (see Table 5.28).

When judged in this way, excluding non-IE-participants, which seems a more accurate method of interpreting the effects of Intervention, the results are not only more satisfying, but are also closer to what happened in the Intervention Exercise itself (see Chapter 5, Section 5.3.4.1). This would also make the overall success of Oral assessment in the AARP look closer to 90%; however, clearly, it is important to try and understand what may have happened in the remaining exceptional case of Group E, previously referred to as a ‘rogue’ group.

6.2.1.2 The rogue group

In Sections 6.1.3.3 and 6.1.3.4 of this chapter, a number of hypotheses were posited concerning the apparent lack of success enjoyed by Group E, related to the Peer-, and Self-assessment of the two Home Writing assignments, such as the size of the group and the constituent age-groups within it. The extremely sharp contrast in Oral assessment behaviour between Groups E and F, in the middle year of the Main Study, 2007-2008, and between Group E and the other nine groups in the AARP, seems to warrant at least some attempt at explanation.

A look at the overall profile of learners in Group E in Chapter 4, Table 4.4, reveals no particular differences with regard to gender, ethnic origin or language ability which could have acted as deviant variables, as suggested in Langan et al. (2005). One possible difference which was revealed by perusal of students’ individual Profile Cards and assessment records for this group, revealed the fact that in the Oral assignment production stage, there was much greater collaboration between students in Group E, as compared with Group F. While there was close to 50% collaboration on oral presentation projects in Group E, with 14 out of 30
presenters working with one or more partners, in the case of Group F, there were no such cases of collaboration and each of the 23 students selected their own particular topics and presented alone.

A study of Oral assessment conducted by Peng (2010) could help to throw some light on this matter of collaboration. In his study, all the oral presentations were group collaborations, rather than individual, and he had the learners apply two types of Peer-assessment, which were group-to-group and within-group. His findings revealed a dramatic difference between peer-assessment grades in these two cases, with average grades rising from 66.6% (group-to-group) to 80.4% (within-group) respectively in the Higher-intermediate group and from 67.5% (group-to-group) to 85.6% (within-group) in the Lower-intermediate group. This led Peng to conclude that within-group Peer assessment was “somewhat influenced by friendship effects” (2010, p. 100).

Although no such distinctions in Peer-assessment procedures were made in the AARP study, where Peer-assessment involved the whole class, and included assessment by peer collaborators, when there were any, it is possible that such distinctions in assessment were informally introduced by the students themselves, as they were given complete freedom as to how they would present – alone, in pairs and in trios – with larger groups discouraged simply due to the complications of identifying and grading individuals. As in the case of Peng’s (2010) study, it could also be that in the AARP study that joint-peer-presenters were much less objective in their peer-assessments of their partners and that such camaraderie also gave them an over-inflated impression of their own performance, leading to a much higher degree of subjectivity in grading, both in the case of peer-assessment and self-assessment.
Another very useful source which can offer information about how the students in Group E were thinking and operating is the Assessment Questionnaires which they completed, the results of which can be seen in the t-analysis of questionnaire responses for 2007-2008, in Chapter 5, Table 5.17, and in Appendices 27a to 27j in Vol. 2 of this thesis. Some very interesting facts come to the fore in their answering of these ten questions in which they perhaps display a greater understanding of themselves and a higher degree of honesty than it seems they employed in the assessment process itself. Since this group behaved unlike any other group, a synopsis of the questionnaire findings for that group is offered in Appendix 24.

It seems clear from these answers that Group E was a very troubled group, with very mixed feelings about the assessment methodology being employed, with many perhaps trying to resist it and fight it with all their might. This is apparent from the student comment from a member of Group E, given in Appendix 27f, for example, who states: “I do not believe that a person can be sufficiently objective with himself.” There seem to be many lessons to be learned from this group, which stands apart from all the others in its assessment performance.

Little (1990, p. 12) makes the point that for some learners, autonomy is the last thing that they want and we have to respect the fact that in some ways this group showed its autonomy by refusing to conform to the demands placed upon them in this system of assessment and this refusal comes across loud and strong. Biggs (1999, p. 61) reminds us that learners are accustomed to a system where control of assessment processes is not usually shared and for that reason students take a defensive stance of “watching their backs”. We have to remember that these learners, like most of us, have been through a system where heteronomy is the norm and where competition and rivalry thrive and are rewarded, and where we are all
dependent on knowledgeable others to tell us how well we have done or to what
degree we have succeeded and we should perhaps be happy that 85-90% of the
learners in this study of learner-centred Oral assessment seemed to be willing to
overcome their pre-conceptions about learning, assessment and the roles of teachers
and learners and made a genuine attempt to learn and to understand themselves
better through learner-centred assignment production and assessment processes.

6.2.1.3 Justifying comments for grades awarded
In the final year of the Main Study, as described in Chapter 4, Section 4.5.3, an
administrative change was made in the way in which students recorded their Peer-
assessment and Self-assessment grades and instead of bundles of small papers being
collated for each individual student’s triangulated assessment, each student had a
booklet, with only their student numbers on the outside, where the assessment
criteria sheet acted as a cover. Inside the booklet was something like a register of
students’ names with several small boxes for the recording of ratings as well as
space for comments on the presentation. Although encouraged to do so, in the first
year of its use, in 2008-2009, very few students made remarks. One very interesting
outcome which could be attributable to the Intervention is the fact that the number
of comments made by students in 2009-2010 increased dramatically, an indication
perhaps that they were now better able to express their evaluation in words as well
as in numbers and were also prepared to make that extra effort. Saito (2008, p. 554)
also finds that compared to his control group, one outcome of rater training in his
experimental group was that learners made more salient comments related to
presentation skills, although the two groups did not differ significantly in other
respects. Although all of these comments have been recorded, it is regrettable that in
the AARP, there was no time to collate these comments, whether positive or negative, and feed them back to the presenters.

In the next section, other studies of learner-centred Oral assessment will be considered and compared to the AARP.

Findings from other oral assessment research studies
As in the case of writing assessment, it is very difficult indeed to make comparisons with other Oral assessment research studies since the whole approach taken in the AARP and the assessment philosophy on which it was based was somewhat unique, not only in the way it was conducted, but in the rate of success that it enjoyed. Other studies offer training for Peer-assessment, and proceed from there to Self-assessment, whereas the AARP uses Peer-assessment itself as the training-ground for Self-assessment.

Other studies (Patri, 2002; Chen 2006a; 2006b; 2008; Saito, 2008; Peng, 2010) break the whole group into smaller working groups which exchange opinions on oral presentation performances and seem to offer moral support when it is that group’s turn to perform. There is constant dialogue and negotiation on rating within groups, between groups, and with the instructor, before committing to rating particular performances, even their own. This philosophy seems to hinge on a process of negotiation to reduce subjectivity and bring about objectivity, while the AARP aims to encourage more of an internal dialogue with reference to the criteria and without the influence and the possible obstruction of the opinion of others, whatever their level of expertise. This is based on the belief that years of other-directed thinking and approval have to be overcome if a move away from heteronomy and towards reflection, ownership of learning, self-directed thinking
and autonomous criterial judgement and decision-making is to be achieved, and in most cases it seems to have been effective.

In all the research literature examined, only Natri (2007) and Chen (2008) make it very explicit that their aims in using learner-centred assessment are to promote autonomy and Natri describes a system of assessment she has put into place for all the skills, based on CEFR levels, and the results derived thereof, rather than one particular study. Chen’s instruments bear similarities to the AARP, but the results she produced are the exact opposite, with I-A means coming highest, followed by P-A and S-A, but she finds that Peer-assessment is comparable to Instructor assessment, unlike Self-assessment. Patri (2002), also finds that Peer-assessment comes closer to Instructor assessment, while Self-assessment is considerably different. Peng (2010, p. 98), as mentioned previously, finds that Peer inter-group assessment is comparable to Instructor-assessment, but finds that Peer intra-group assessment shows clear signs of what he refers to as “friendship effects”.

Generally, there are many reasons to be pleased with the results from AARP Oral assessment.

6.3 Answers to the AARP research questions
Taking into account all the findings from the AARP, and the evidence and support which has been taken from other research studies to offer some explanation for some of those findings, it should now be possible to verify if the four research questions (see Chapter 1, Section 1.4), on which the hypothesis of this thesis is based, have been satisfactorily answered.
The 1st research question
Concerning the first question: (1) if learners could assess the oral and writing skills of their peers, according to predetermined criteria, with objectivity and reliability, it seems to be almost indisputable that given the type of peer-assessment put into practice with oral skills in the AARP, that learners were able to adopt the criteria for assessment and use them effectively and in most cases did so with a high degree of objectivity and reliability. In the case of writing also, there are strong indications that learners took the task of peer-assessment seriously and were, for the most part, following the same frame of reference as the instructor with the criteria. The degree of responsibility which was involved in peer-assessment of writing, since it involved a single peer-assessor, was clearly much greater and therefore more challenging than was the case with assessing speaking. There was also greater difficulty in the 2nd writing assessment task compared with the 1st task; nevertheless, in most cases, students made a brave attempt and there are strong indications of this in the rating alignment between P-A and I-A (see Chapter 5, Table 5.31 and Figure 5.2).

The 2nd research question
With regard to the second research question: (2) if learners, using peer-assessment as a stepping-stone, could then assess their own oral and writing skills performance with the same objectivity and reliability, using the same predetermined criteria, while it is clear from the AARP and from other research projects that self-assessment, whether of oral skills or writing skills, is, generally, a much more challenging task, unlike Self-assessors of oral skills from Taiwan (Chen, 2008; Peng, 2010) and Japan (Saito, 2008), the students in the AARP marked themselves consistently higher than Peer-assessors and the Instructor, yet, for the most part, not to an unreasonable degree. Correlation coefficients reveal a genuine attempt to use
the given criteria in the same way as the Instructor. There are clear indications from this research project that the process of Peer-assessment of presentations helped students to curb their tendencies towards self-inflation and that for the most part students did try to exercise restraint and objectivity.

In the case of writing, it is possible that more attempts at peer-assessment, though much more complicated to organise, might have been beneficial in offering a variety of samples to students for comparison; nevertheless, in the form the AARP took, the Pre-Study stands as a close-to-perfect example of what could be achieved and it is possible that if in later years the groups had also been more distanced alphabetically, that greater objectivity could have been exercised both with Peer-assessment and with Self-assessment. In the studies that were reviewed, there were no examples where Self-assessment was deemed altogether reliable, so what was achieved in the AARP overall in terms of Self-assessment, could be considered a moderate success. It seems clear, in comparing with other studies, that both in the cases of over-marking and under-marking that psychological and affective factors have come to the fore and that such factors, according to researchers such as Kohonen (1999), AlFallay (2004) and Chen (2006a), are very hard to deal with.

The 3rd research question
Concerning the 3rd research question: (3) whether there is evidence that learners assumed ownership of the assessment criteria checklists and exercised judgement in an atmosphere of cooperation and trust, with regard to Oral assessment processes, with the exception of one group, Group E, and also some members of Group I, who did not take part in Intervention processes, there would seem to be more than sufficient evidence that students responded very well, both to the challenge of executing their presentations and to the challenge of assessing them, both in the
capacity as Peer-assessors and as Self-assessors. Although not involved in the creation of the criteria check-lists for the assessment of the oral assignment, students willingly ascribed to them and by foregoing, for the most part, the opportunity to manipulate the situation and their “position of power” (Ritter, 1998, p. 79), used the criteria to honestly and critically assess their peers and themselves, in a manner resembling the instructor.

In the case of Writing skills, there is strong evidence that learners tried very hard to rate and mark assignments according to the criteria given and did their best to overcome any feelings of anxiety or concern about final course grades and course outcomes. Though clearly challenged by the greater complexity of the 2nd Writing assignment, both in its production and its assessment, students for the most part rose to the challenge and did their best, assuming greater responsibility for assessment and learning processes. Evidence of close alignment between P-A and I-A in the 2nd Home Writing assignment, through P:I frequency ratios have already been referred to and are outlined in detail in Chapter 5 (see Section 5.4.2).

The 4th research question
Regarding the final research question: (4) whether there was evidence that the cyclical process of Peer-, Self- and Instructor assessment had afforded opportunities for greater self-direction and autonomy, it seems clear that learners came to develop much greater awareness of themselves as learners, firstly through awareness-raising exercises (see Appendices 1 & 2) and consequently through assessment processes (see student comments in Appendices 26, 27 & 28), and to have a much clearer picture of their strengths and weaknesses as language learners, both through the Learner/Teacher Contracts (see Appendix 4) and on account of assessment processes (see Appendices 26, 27 & 28).
By their involvement in cooperative assessment processes, which involved assessment of learning, which became assessment for learning and assessment as learning (Stiggins & Chappuis, 2006), learners became less dependent on spoon-feeding and were required to use critical analysis, which resulted in deep, rather than superficial learning. There are clear indications, both quantitative and qualitative, that change was effected by involvement in the AARP and that processes were set in motion for learners to move further along the continuum of autonomy (see the AARP model for autonomy in Table 2.8).

Chen (2008, p. 253) postulates that her students, through their participation in self-assessment, were, based on Littlewood’s (1999, p. 75) model, able to adopt a more “pro-active” role in their learning, a clear indication of, at the very least, “reactive” autonomy in practice (see Chapter 2, Figure 2.4).

In the AARP model for autonomy offered in Chapter 2 (see Table 2.8), it was emphasised that the degree of autonomy exercised can vary from individual to individual, from task to task and even from moment to moment, but the evidence presented here seems to show that learner-centred assessment, if used efficiently, can help maximise the learning opportunities offered in the language classroom and lead the learners towards greater autonomy.

van Lier (1996, p. 13) asserts that given that adequate provision of learning opportunities is offered in the educational setting, it is the autonomous learner who is “responsible for learning as well as lack of learning”, while Sinclair (2009, p. 185) reminds us that “learner autonomy is a construct of capacity which is operationalised when willingness is present”. The importance of “willingness” is also emphasised by Littlewood (1996, p. 427; 1997, p. 82).

Clearly, the degree of uptake depends on the learners’ predisposition and willingness to accept both responsibility and change; nevertheless, the hypothesis on
which the thesis is based, that there are degrees of autonomy and that it can be promoted through peer- and self-assessment, seems, for the most part, to have been verified.

### 6.4 Concluding remarks

In this chapter, the outcomes of the AARP have been discussed at length and comparisons have been drawn between this study and other projects with similar aims. The outcomes of writing skills assessment and special cases and phenomena which emerged from the AARP were highlighted and discussed and compared with the findings of other research studies. Likewise, the findings derived from oral skills assessment in the AARP were discussed and particular cases of interest were elaborated upon and compared with examples from the literature. Finally, responses to the four research questions were expounded upon and conclusions concerning the underlying research hypothesis were reached.

Chapter 7, the final chapter, which follows, will examine if the aims and purpose of the thesis have been met, to what extent the research was successful and any limitations which were discovered. In addition, any areas deserving of further research in the future, will be outlined.
Chapter 7

Conclusion

7.0 Introduction

This concluding chapter aims to summarise whether the aims of the Assessment for Autonomy Research Project (AARP) were met and to outline in what ways the AARP was successful and in what ways less successful. Consideration will be given to what the AARP has contributed to knowledge in the areas of autonomy and assessment in EFL and concerning the relationship between them. Possible limitations of the study and ways in which the research could be moved forward in the future, will be suggested.

7.1 The aims of the AARP and whether they were achieved

The AARP was an attempt to investigate the involvement of students of English Philology in the School of English, Aristotle University of Thessaloniki in evaluation of their (EFL) language learning, and by assimilating and analysing both the products and the processes of this involvement in Peer-assessment and Self-assessment, to estimate its value in promoting critical thinking, reflection, self-direction and, consequently, greater degrees of autonomy in their language learning.

The basic hypothesis was that there are degrees of autonomy in language learning which can be promoted by greater learner involvement in assessment, and this was to some extent proved correct, as shown by the results in Chapter 5 and the discussion of the results in Chapter 6, particularly in the case of Oral assessment. The two preliminary research questions which involved the finding of a suitable definition of autonomy to suit the conditions of the AARP as well as the finding of a suitable model for the AARP seemed to be satisfactorily resolved by selecting a “glocalised” (Schmenk, 2005) definition for autonomy, which would at one and the same time
meet universally agreed principles concerning autonomy and, equally well, suit the local setting and the nature of the participants. Also, by demonstrating the usefulness of viewing autonomy as being on a continuum with heteronomy at the other end of the spectrum\textsuperscript{34}, and by showing how various theories of learning and autonomy in language learning actually converge and fit into that continuum, it was possible to build on that model and add into it Little’s (1996a; 1999a) four key questions which equate with levels of understanding, necessary for autonomy, of What? How? Why? and With what success? Several existing models of autonomy were examined as to their suitability, but finally all were rejected in favour of a modification of a model for lifelong learning devised by Stolk \textit{et al.} (2007), which accommodated the AARP idea of degrees of autonomy on a cline with heteronomy, as well as allowing for fluctuation, progression, and even regression on the various levels and degrees. With this adapted model (see Chapter 2, Table 2.8), the second preliminary research question, concerning finding a suitable model for the AARP, also seemed to be satisfied.

In addition, the four main research questions also seemed to be proved to a large extent by the AARP results, so that: (a) learners for the main part were found to be able to assess the oral and writing skills of their peers, according to predetermined criteria, with objectivity and reliability; (b) using Peer-assessment as a stepping-stone, learners seemed to be able to assess their own oral skills performance (Self-assessment), with a reasonable degree of objectivity and a reasonable degree of reliability, and to assess their own writing skills performance (Self-assessment) with a slightly lesser degree of objectivity and a slightly lesser degree of reliability, for the most part, using the same predetermined criteria; (c) there was significant evidence

\textsuperscript{34} This continuum, as has been made clear from the many pedagogical/theoretical continua presented, is not purely linear in the sense condemned by Schmenk (2006), but is, rather, multidimensional and multifaceted and allows from a great deal of fluctuation and variation.
that learners assumed ownership of the assessment criteria checklists and exercised judgement, with perhaps some exceptions, in an atmosphere of cooperation and trust; (d) there was evidence that the cyclical process of Peer-, Self- and Instructor-assessment could combat pre-conceptions about assessment, in a large number of cases, and turn assessment into a learning opportunity, through the development of criterial thinking, reflection, decision-making, ownership and self-determination, thus promoting autonomy.

7.2 The success of the AARP

The unique value, or interest, that lies in the AARP, compared with existing research, concerns not just its scope, in trying to establish an essential link between assessment and autonomy, but in its length, extending over a five-year period, and in its depth, in that it investigated triangulated assessment of both the writing and speaking skills of the participants involved. Not all the research studies which deal with assessment of specific writing and speaking tasks have been conducted with the aim of investigating autonomy, particularly in an HE EFL setting, with only Chen (2006b; 2008) elaborating on its promotion in some detail. The AARP therefore goes some way towards overcoming this possible oversight.

In addition, all of the assessment research studies, whether on writing skills or speaking skills have had the duration of one semester only. The AARP also had the duration of one semester, but was repeated over five consecutive autumn semesters. Also, with one exception (Cheng & Warren, 1999; 2005), all of the studies either dealt with assessment of speaking skills or, to a lesser extent, of writing skills, but did not deal with assessment of both speaking and writing tasks performed by the same students during the same semester. The richness of the AARP data, which is
concerned with the two productive skills of Speaking and Writing, permits some interesting comparisons to be made.

### 7.2.1 Replication of research conditions

One particular advantage of the AARP is that the same pedagogical, methodological and evaluative procedures were replicated, as far as possible, year on year. This attempt at replication seems to be important since Boud & Falchikov (1989, p. 535) note, with regret, the evident absence of such replication from the assessment literature, and yet more than 20 years later, after their complaint was voiced, this gap in the literature does not seem to have been filled. The AARP could be the first of many replicative studies which might contribute to filling that gap.

Indeed, the great lack of consistency between studies, both in the way they were conducted and in the way that they were analysed, mentioned by Boud & Falchikov (ibid), Saito (2008) and Ross (1998), continues to this day, rendering it extremely difficult to make comparisons and draw verifiable conclusions, so studies of longer duration which seek to replicate, such as the AARP, may help to reveal more about the particularities of assessment behaviour in a higher education EFL context.

### 7.2.2 The missing autonomy link

While theoreticians in the EFL world such as Holec, Oscarson, Little, Dickinson and Kohonen, whose work is reviewed in Chapter 3, Section 3.2, have taken great pains to point out the essential link between autonomy and assessment, these strong ties have not been so apparent nor been given so much prominence in the HE research studies on Peer-assessment and Self-assessment which have been conducted by more recent researchers. With a few exceptions, such as Chen (2006b; 2008), Natri (2007) and Stickler et al. (1999), although the connection between learner-centred assessment and more effective learning has been made, the important role that assessment can
play in the development of learner autonomy in language learning, particularly in an EFL HE setting, has prompted no more than passing reference and, in some cases, no reference at all.

Indeed, it is not always very clear whether researchers who aim to promote Peer-assessment and Self-assessment in their classes fail to mention the important role of cooperative assessment in fostering autonomy through neglect of this important aspect or through fear that they have found insufficient evidence to support such a theory. Otherwise it is difficult to understand why some researchers have invested so much time and trouble on training for assessment, particularly with students who are not majoring in FLs, except if they believed, quite reasonably, that such skills could be transferred to other subjects and disciplines.

A possible stumbling-block for some language instructors is the fact that autonomy, being very much an abstract concept rather than a concrete actuality, is very difficult to pinpoint and measure, especially since it is not something static or constant or something which necessarily grows with advancement in language learning (Kumaravadivelu, 2003, p. 144). Clearly, more research in the area of assessment and its link with autonomy has to be published in order to engender interest and encourage further exploration and investigation on the part of language instructors.

7.2.3 Cross-sectional and longitudinal perspectives
What could be considered another advantage of the AARP was the use of two research groups in each year and the formulation of three distinct stages in the research study, with a Pre-Study in the first year, a Main Study, with a duration of three years, and a Post-Study of one year, which made for a very interesting ‘weaving’ process in the research project, with cross-sectional comparison made
possible within years, the ‘warp’, and also providing longitudinal sections for comparison over the years, the ‘weft’. This enabled possible patterns and anomalies in the fabric of the project to be identified and subjected to closer examination.

However, what could be considered a disadvantage of the AARP, was the fact that, like all the other published studies of this type, the course on which it was conducted was of short duration, from 10-13 weeks, and so it was not possible to study any long-term effects or outcomes concerning particular individuals or groups of learners.

7.3 Usefulness and transferability of the research
It is hoped that the AARP will be useful in adding to the existing body of knowledge related to the use of learner-centred assessment practices in EFL HE, and the use of such practices to promote autonomy, but its usefulness will extend still further if some of the theory and practice behind it can be taken and applied in other institutions and in other countries.

7.3.1 The AARP model
Bearing in mind the previously-mentioned fact that the link between assessment and autonomy has not been given the full attention it deserves in EFL research, one way in which the AARP may be useful to other researchers is in the model for autonomy proposed in Chapter 2 (see Table 2.8), based on Stolk et al. (2007), which takes into account the important guiding questions, for what Little (1996a; 1999a) sees as the four necessary levels of understanding for the promotion of autonomy in learners, of: (1) what are we learning? (2) why are we learning? (3) how are we learning? and (4) with what degree of success?, which seems to work successfully in the present context since it places emphasis on the importance of assessment to learning and of a parallel dual progression towards greater autonomy, through greater learner-centred
assessment, as envisaged also in the continuum (see Chapter 2, Fig. 2.11) created by Harris & Bell (1990, p. 111). The model proposed in the AARP seems to be a valid one in that as well as allowing for progression along a continuum, it permits of both regression and fluctuation. The same model could equally well apply to other levels of education, to other cultural settings, and to more technologically-oriented approaches such as blended and on-line learning environments.

7.3.2 The AARP ‘glocalised’ definition of autonomy
Another way in which the AARP might prove useful is in inspiring more “glocalised” thinking in relation to autonomy in language learning. The “glocalised” definition of autonomy which was proffered in Chapter 2 (see Table 2.6), influenced by the thinking of Schmenk (2005), aims to show that autonomy may not necessarily mean the same in all instances, and rather than following some universal ‘ideal’ of autonomy it is possible to adapt the definition to suit our own teaching circumstances and learning environment, to suit our own “glocalised” and practicable form of autonomy. Key to the idea of autonomy in the AARP is reflection and decision-making, and decision-making in this instance is promoted through Peer- and Self-assessment, with Peer-assessment acting as a training-ground or stepping-stone for Self-assessment. Carless, Joughin, Liu and Associates (2006, p. 12) believe that by involving learners in assessment, learners “develop the capacity to evaluate their own work”, which is regarded as an essential element in lifelong learning.

7.3.3 The AARP quantitative and qualitative data
In a previous section, the bulk of the data produced by the AARP is mentioned and one way in which the AARP makes a useful contribution to the EFL literature and can inform future teaching and assessment practices is through the large amount of both quantitative and qualitative data which was gathered over the five-year period.
While the quantitative data was based on 2 writing assignments and 1 speaking assignment, the qualitative data was drawn from responses to questionnaires, which were collated and analysed. It is perhaps in the findings of the qualitative data that the similarities between this study and other studies are most apparent. The qualitative data has been analysed quantitatively and a very small sample of student statements has been linked to the responses to questions illustrated in the xyz graphs in Vol. 2 (see Appendices 26a-26j, 27a-27j and 28a-28v). Such statements are extremely enlightening and could form the basis of a future comparative study.

7.3.4 Two types of peer assessment
Apart from focussing on the two skills of speaking and writing, which to some extent sets the AARP apart from other EFL assessment studies, which, as previously mentioned, with the exception of Cheng & Warren (1999; 2005), tend to focus on the assessment of a single skill, the other aspect to this study, which makes it special, is the fact that two types of peer assessment were involved: whole-class peer assessment in the case of speaking, and single peer-assessment in the case of writing. While these seemed like altogether natural procedures on the AARP, other studies have tended to take quite a different approach, both in the assessment of speaking skills and in the assessment of writing skills.

In her analysis of group peer assessment, Falchikov (2005, p. 201) talks about “multiple raters” being more reliable than “individuals” and she highlights the use of intra-group and inter-group Peer assessment in some studies with their concomitant use of feedback in assigning grades. She also talks at length about the various forms of bias which can affect Peer-assessment and the various difficulties which can arise through collaborative peer task completion and its assessment, but she does not seem to mention whole-class Peer-assessment of the type that was utilised in the AARP for
speaking skills, in her literature review, possibly because use of whole-class Peer assessment of this type has not been reported to date, or certainly, it seems, not within the EFL literature.

Information concerning one-to-one peer rating of written assignments is similarly lacking, since, in most studies, students are given three or more peer assignments to assess at home (Matsuno, 2007; 2009; Saito, 2008; Saito & Fujita, 2004). Although such a system would have been exceptionally difficult to implement and even more difficult to collate in the AARP, it might, nevertheless, have been a good opportunity to allow students to see a range of student writing, particularly as the research groups in the AARP in each case were mixed-ability classes. One apparent danger with assessing the writing of only one peer is that students may have remained with a rather jaundiced or biased view after assessing only one example of either exceptionally good or exceptionally poor writing, since the only other piece of writing they had to compare with was their own. Indeed, it would be interesting to have access to studies closer to the AARP in their use of Peer-assessment in order for comparisons to be made, particularly in the case of oral assessment where the assessment system put into practice in the AARP appears to have been particularly successful.

7.3.5 Use of a single instructor-rater in the AARP
Another obstacle to comparison with other EFL studies is the fact that, in some cases, the number of instructor-raters involved in each instance is not made very clear, or that with the exception of Matsuno (2007; 2009) in her Pre-Study assessment of essay-writing, Chen (2006a; 2006b; 2008) and Peng (2010) in their assessment of oral presentations, researchers opt to use ‘mystery’35 multiple instructor-raters both for

35 The word ‘mystery’ is chosen deliberately here, since these raters are unseen and unknown to the student participants, and, like ‘mystery shoppers’ are employed as a form of quality control.
speaking and writing assessment. Again, in the case of the AARP, the instructor-researcher was, as an LM I course instructor, in any case required to assess her students single-handedly and so took the decision to remain a ‘lone rater’, fully conscious of the dangers in the assumption that “the tutor mark is always correct” (Orsmond et al., 1997, p. 358). Since the aim was to encourage learner autonomy, it seemed more logical to “move towards a more democratic approach” to assessment (Somervell, 1993, p. 229) by involving the learners themselves in the assessment process and focussing more on the “validity of judgments” rather than the “reliability of grading” (Sadler, 1989, p. 122). This decision to remain a lone rater seemed altogether natural on the AARP, particularly in view of triangulation of the I-R’s scores with those of self-raters and peer-raters.

Although accepting Alderson’s (1994, p. 13) view, that “nobody is free of bias”, and that where “judgement is involved, however well informed, it is not objective” (Alderson, 1994, p. 4), as a fully-trained, fully-standardised and fully-qualified international oral examiner for Cambridge ESOL, of many years’ standing, the I-R thought herself capable of unbiased marking in the oral skills, but at the same time found Chen’s (2008) idea of recording all the presentations and rating them independently for a second time, worthy of emulation, even if not practical and practicable in the case of the AARP, while the idea of employing a panel of instructors to rate from the video-recordings of presentations (Saito & Fujita, 2004), or to rate batches of student writing assignments (Matsuno 2007; 2009) seemed somewhat surreal, rather than objective, and likely to cause more problems to surface, with regard to standardisation and agreement, than to dissipate.

Although the reasons for using external ‘panels’ of instructors against which to draw comparisons, in both studies of speaking skills and studies of writing skills, are never clearly explained nor justified, it seems to indicate more conservative views of
TEA and an attempt to maintain some form of fairness and objectivity, which is neither explained nor voiced; nevertheless, a close reading of these studies (Matsuno, 2007) reveals that in the case of writing in particular, there were tendencies for panel members to disagree, and longer training sessions had to be introduced in an effort to combat rater disagreement among the ‘mystery’ rater-instructors. Clearly, the use of multiple instructor-raters can create more problems than it solves and, as previously stated, it would be useful to have more studies using a lone, but experienced, instructor-rater against which to compare experiences on the AARP.

7.3.6 The simplicity of the AARP
One more way in which the AARP might be of use to other researchers and be worthy of emulation is in its simplicity. In the criteria check-lists, by using 5 criteria on a Likert scale of 1 to 5, the assessment system was simple, without being oversimplistic. Other researchers seemed to lack confidence in creating their own criteria check-lists and based theirs on lists created by others, which tended to be long and had not necessarily proved successful in the past. They seemed to forget that these lists and scales would be used by inexperienced and “naïve” raters (Jafarpur, 1991), who had to be able to quickly assimilate both the criteria and the rating-scale.

Other researchers insisted on using a scale of 1 to 4 or 1 to 6, in order to avoid having an average value (3, in the case of a rating scale of 1-5), which they believed that learners would be tempted to over-use, or in an attempt to encourage use of lower-end rating values. Choosing the scales that they did seemed to add to, rather than diminish, the problems in the already difficult task of learner-assessment. The fears of the above-mentioned researchers, in this respect, are therefore worthy of further investigation and this topic in itself would make an interesting study. The only evidence of over-use of the mid-point in the 5-point rating scale in the case of the
AARP, was perhaps in responses to the AARP questionnaires rather than in rating of assignments. Overall, the checklists and the rating scale seemed to work well in the AARP.

7.4 Limitations of the AARP

While there were many elements in the AARP which seemed to work successfully and might be worthy of emulation; nevertheless, like all research studies there were elements which could have been given more emphasis and prominence and perhaps been more clearly formalised, in order to make them more transparent to the participants and to other researchers. One such element is that of feedback, a subject which was not given a great deal of emphasis in the EFL research studies which were scrutinised, except for its use as a form of interactivity in peer-rating, while it holds much more interest in the use of evaluation as “a prelude to action” (Alderson 1994, p. 10).

7.4.1 Assessment feedback on the AARP

Essentially, there are two types of feedback which students require when they are involved in learner-centred assessment. One concerns information regarding their performance on the actual writing or speaking tasks and since the assessment of each task is triangulated, there is feedback from the peer-assessor as well as the instructor-assessor which has to be fed back (see Chapter 4, Figure 4.2), with a view to the writers or presenters involved being able to take note and apply the information in future tasks, so that feedback becomes feedforward (Hounsell, 2007; Knight, 2006; Murlagh & Baker, 2009).

With regard to the oral presentations, unfortunately, this kind of feedback was lacking on the AARP, basically because it was a one-off task and generally came at the end of the semester, and also because written peer feedback was lacking on oral
assessment until the last year of the study (see comments on this in Chapter 6, Section 6.2.1.3). While the I-R kept notes and recorded comments on presentations, it was very unusual for students to ask the instructor’s opinion, perhaps because their main interest was in the grade. While there is no guarantee that students would have paid attention to the instructor’s comments if more formally recorded and handed over, on the other hand, it may have been a lost opportunity to change their cultural thinking in terms of assessment, and perhaps be persuaded that comments were more valuable than grades. Perrenoud (1998, p. 87) takes the pessimistic view that feedback is like messages in bottles “thrown out to sea”, with no guarantee that “the message they contain will one day find a receiver”.

With regard to writing, in the first three years of the study, students were offered photocopies of corrected assignments, while in the last two years all three corrected copies, produced by Self, Peer and Instructor, were returned to students for their perusal and safeguarding. Any questions which emerged from this could then be addressed to the instructor.

The other type of feedback, which the instructor took great pains to feed back to students in a face-to-face encounter, usually with one or more classmates in a non-threatening atmosphere, was information concerning their Peer-assessment performance with regard to the two writing assignments and their Self-assessment performance with regard to their two writing assignments and their oral assignment. In other words, it was important that since learners were being employed as raters and assessors, and particularly since their involvement contributed to actual task grades, that they should be informed and made aware of how close or how deviant their assessment was from that of the instructor and their peers. This was a vital opportunity for students to verbally account for the grades which they had awarded.
The importance of both types of feedback cannot be over-stressed and yet it does not feature largely in the literature. A type of feedback which did feature in other studies and might have been emulated in the AARP was the use of intra-group and inter-group discussion and feedback as a way to standardise assessment of both speaking and writing. This approach could have been adopted in the AARP, not in the case of ‘real-life’ or ‘live’ assessment, which to be truly objective the instructor-researcher still believes must be conducted alone, in order to achieve “interaction with oneself” or the “detachment” Little (1990, p. 10) sees as necessary for autonomy, but could have been used as a form of mock assessment, which would offer valuable practice, with support. Sharing and exchanging views of different written or spoken language performances could have been beneficial in boosting confidence in their ability to assess and would have familiarised students with the assessment criteria, but in a relatively controlled environment and on a small scale.

Feedback of the two types just mentioned could have been given greater prominence in the AARP and certainly needs to be reported in much more detail by other researchers, particularly with regard to the aspects of speed and immediacy, which seem to be vital if feedback is to have any effect and any usefulness at all. Feedback is of no value if learners have already forgotten details of what production of the assignment entailed.

A third type of feedback which is missing, for the most part, from the literature, and could be very useful, is feedback from the researchers themselves about how this approach to assessment changed their approach to teaching and how, in spite of an increased workload, and, possibly, other repercussions, perhaps from their institution or colleagues, they remain convinced of its benefits. Two very realistic accounts, in this respect, are offered by Stickler et al. (1999) and Natri (2007). More accounts of this kind would be very useful as would more detailed accounts from researchers
concerning the exact procedures that were followed in their studies related to who did what, how they did that, when they did that, how long it took and, in particular, why they did that, thus providing information which goes beyond just the results. It is the fine detail in assessment procedures which can be of tantamount interest and importance.

7.4.2 Individual assessment behaviour
Related to the matter of feedback, just dealt with, is that of individual assessment behaviour as opposed to group assessment behaviour. In this study, as in most research studies, the focus has been on group assessment behaviour rather than individual assessment behaviour, so that it was not possible, as part of this research to monitor particular individuals, beyond the cross-sectional ‘eye-balling’ (McNamara, 1996) of data permitted by the Learner Profile Cards. Application of a particular form of statistical analysis, like Rasch Facets Analysis (see Chapter 3, Footnote 14), had it been available, could perhaps have been used not only to monitor the assessment behaviour of individuals, but could have been used as a form of ‘intelligence’ to convince individuals of their assessment accuracy or lack thereof, with technological backup, perhaps convincing them of the importance of their assessment contributions and causing them to take the process more seriously and assess with greater care and accuracy. Matsuno (2007) emphasises the usefulness of Rasch Facets Analysis in this respect, but discloses no details as to how information of this kind was fed back to students, if indeed it was. Matsuno states that she was awarding marks to the students for their assessments, but does not clarify if Rasch Analysis was being used to help her with the awarding of grades or if it was being used as a mechanism to threaten the students in some way that their assessments were somehow being checked and “policed” (Tan, 2004).
In the case of the AARP, individual assessment behaviour certainly could have been examined more closely in the Post-Study, where more evidence of individual behaviour was gathered in the Intervention exercises. Closer scrutiny of one anonymous example of Writing assessment (see Appendices 14a & 14b) reveals close coincidence between the Anonymous Assessor and the I-R in rating in the IE, but, paradoxically, considerably less agreement when it comes to the same Assessor’s actual Peer-rating of the same 1st Home Writing assignment. The difference in alignment between this learner’s mock Peer-assessment and actual Peer-assessment and Instructor-assessment of the same can only be explained by affective factors coming into play.

7.4.3 An examination of outliers
In part, because of the emphasis on whole-group assessment as opposed to individual assessment, there was no real attempt to “police” (Tan, 2004) the assessment of individual students on the AARP. Policing and incriminating accusations were generally avoided in order to create an atmosphere of trust. This means that there was no ‘witchhunting’ or attempt to identify outliers, whether they were overestimating or underestimating, and no attempt to remove them from the data. In addition, with the exception of Group E, which was the true ‘rogue’ group in the AARP, there has been no real investigation, in this study, of particular outliers and whether there were questions of age, gender or ethnic origin which could have played a role either in their successful or deviant assessment behaviour. This would make an interesting further study, as would students’ scoring of particular criteria as compared with the instructor (Everhard, unpublished data). Examples of such comparisons are provided for one anonymous participant in the Post-Study Writing IE (see Appendices 14a & 14b) and for the Post-Study Oral IE (see Appendix 16).
7.4.4 Lack of longterm development

Another, equally important failing of the AARP, previously referred to in this chapter, is that it was not possible to follow the same individuals or groups through different stages of learning. Indeed, some of the students commented that they did not yet feel capable of assessing their peers and themselves accurately, while they thought they might be more capable of doing so in their later undergraduate years. The question of “maturity” has already been mentioned in relation to assessment, and it would be interesting to examine if age is a determining element and what other elements exactly contribute to a person’s assessment maturity.

7.5 Conclusions and recommendations

Kumaravadivelu (2003) reminds us that in order for there to be any hope of promoting autonomy in the language classroom, there are two basic requirements, which are: (1) support from within the institution, and (2) an appropriate classroom culture. In the first instance, ‘support’ within an institution has to go beyond the point of simply not causing barriers or obstructions, but rather has to be more of a philosophy (Nunan, 1997) or ethos (Everhard, 1995) to which everyone in the institution subscribes, from the humblest to the most powerful levels, and includes both administrative and academic staff. Students will only become confused if some instructors encourage their autonomous growth while others wish to retain their control over learners and do everything, albeit unconsciously, to stunt that growth. Academics have to recognise that the form of evaluation and the examinations they set have a powerful influence over the way that students learn and have to realise that examinations which are easy to administer and correct are not necessarily promoting the kind of deep learning that they would like to see their learners undertaking.

Administrators also have to be aware that grouping and timetabling of students can put learners both at an advantage or a disadvantage, with regard to preferred
learning styles, and that elements of choice and flexibility in programming can encourage responsibility and growth. In addition, those who implement Peer- and Self-assessment also have to realise that this form of assessment is equally pointless if it is used only to generate grades. It then becomes just another form of summative assessment and important learning opportunities are lost. Preparation of learners and teachers for autonomy through assessment, with its underlying recalibration of power and fostering of criterial thinking and assessment literacy is therefore of the utmost importance (Everhard 2011a; 2011b).

7.5.1 Preparation of learners
A significant element in the AARP was the preparation, through awareness-raising exercises, which were used in the first weeks of the Language Mastery I course (see Appendices 1 & 2) to try and help learners in their transition from secondary education to tertiary education, and in the move from a system of spoon-feeding (McKay & Kember, 1997) to one where, in order to succeed, they must have learned how to learn and where they are required to use research skills and critical thinking skills and to be able to produce original work. The majority have never really reflected upon themselves as learners or as language-learners and for years have simply been imbibing pre-determined and pre-packaged pieces of knowledge, which they have not learned to connect in any way. This is what makes learner-centred assessment particularly difficult for them. As with maturity, the role of academic literacy in assessment deserves further exploration, and the ways in which learners can be ‘prepared’ to accept this approach to assessment, which informs learning and promotes autonomy, and offers lifelong learning skills, has to be better understood.
7.5.2 Preparation of teachers

It is widely accepted that language teaching has entered what is being referred to as the post-method era (Kumaravadivelu 1994; 2001; 2003). This has implications both for the way that languages are taught and the way that they are assessed and tested. If, as teachers, we support the idea of autonomy, then this has to be reflected not just in the way that we teach, but also in the way that we test, evaluate and assess. This is difficult for teachers for two reasons. Firstly, for the most part, teachers themselves are the products of educational systems which were more supportive of heteronomy, and the risks and paradoxes involved in moving in the opposite direction, are all too apparent. Secondly, the training of teachers has not always prepared them to take account of new approaches to assessment (Mok, 2011).

One of the paradoxes is, of course, that it becomes more difficult to use a course textbook (Fenner, 2000, p. 80; Little, 2000a, p.41-42) as the textbook itself fosters heteronomy, with the coursebook writer determining what will be learned next rather than the teacher or, more importantly, the learners. Without the textbook as the locus of control, there is a danger that both teachers and learners will become disorientated, so clearly both have to be prepared to follow a new, never-before-explored path; however, the rewards are that following such a path gives Peer-assessment and Self-assessment a whole new significance and meaning and learners themselves will be more accountable for their learning. Teachers, however, have to be able to accommodate and live comfortably with “a more complex and less predictable environment” (Nolen, 1995, p. 207), as well as “unpredictable outcomes” (Fenner, 2000, p. 85; Allwright & Hanks, 2009, p. 71).

Little (1995a, p. 178) makes strong arguments for teacher autonomy as well as learner autonomy and reminds us of the fact that though bound by the curriculum, “the teacher cannot help but teach ‘herself’”, so that what she communicates to her
learners is “her necessarily unique interpretation” of the curriculum. Likewise, Sougari and Sifakis (2007, p. 205) highlight the fact that teachers “follow a norm-bound approach” to teaching, driven by their “own deeply-rooted language learning experience”, while Lortie’s “apprenticeship of observation” (1975, cited in Sougari, 2011) is also well-documented. From that point of view, it could be argued that the AARP had a small but important role to play in shaping the opinions of future generations of teachers, since that was the career most of the AARP participants were destined to pursue.

7.5.3 Use of technology
Technologies which support and exploit Peer- and Self-assessment are already being used in on-line learning settings (Murphy, 2008; 2011), with particular success. Learning and assessment of that learning which can be conducted anywhere, anytime, on-line, holds particular appeal and Cambridge ESOL, for example, already makes use of a combination of face-to-face and on-line training, for its oral examiners (OEs), which offers the traditional training with video-clips, with plenty of feedback and guidance so that less experienced examiners/raters can work through more materials and practice sessions, with the necessary support, while the more experienced and competent can go straight through to rating and marks collection. The face-to-face training session with a Team Leader (TL), at the same time, provides opportunities for discussion and exchange of views with the TL and the other OEs, which is so necessary.

Although learners involved in peer- and self-assessment are not being trained as professional examiners, something similar, appropriate to their own level and related to the course tasks they have to complete would be very useful, so learners could choose for themselves, how much training, if any, they need to undertake.
Also, a way in which the AARP could be improved upon in the future, to avoid untoward bias creeping in, would be to ensure electronic submission of assignments and make use of technologies such as Blackboard, which might facilitate random selection of assignments for Peer-assessment purposes, allow anonymous submission of ratings, comments and feedback and make the relevant feedback available to individuals, together with their self-assessment and the instructor’s assessment, for comparison. A system which maintains anonymity and confidentiality could be a great boon and inspire greater confidence in the raters. Electronic systems of this type have been developed in other disciplines, such as Hospitality (Lugosi, 2009), where a system called CASPAR (Computer Assisted Self and Peer Assessment Ratings) has been developed, though, as might be expected, it is not without its faults at this stage.

7.6 Final concluding remarks
The AARP followed one of “many different paths” to autonomy possible (Natri (2007, p. 109), in which the aim was to see if something like assessment, which can so often put obstacles in the way of learners and act as a constraint both to their learning and their exercise of autonomy, could, if made more learner-centred, be turned into an affordance. It seems clear that the learning opportunities provided on Language Mastery I through the production and assessment of writing and speaking tasks greatly assisted the majority of students in their transition from secondary education to tertiary education and in their progression from heteronomy to autonomy. While this path and this passage was not altogether smooth, it seems clear that both peer-assessment and self-assessment provided the very necessary stepping-stones (Falchikov, 2007; Weeden, Winter & Broadfoot, 2002), leading to greater self-direction and self-determination in their learning, and helped them to acquire skills which will be very useful to them not only in their continuing mastery of the English
language, but in coping with the many challenges with which life is still to present them.

Nolen (1995, p. 213) states that if, as instructors, we wish to promote autonomy, then it is up to us to create the type of environment in which our learners “can learn to be autonomous by being autonomous”. Little (2000c, p. 46) also reminds us that the “acceptance of responsibility” required for the achievement of autonomy is not a “single act”, but rather a “gradually developing state of mind”, while Allwright (1990, p. 70) points out that the development of autonomy is likely to “take a long time” since “it involves the development of abilities hitherto neglected by educational systems”. Sambell et al. (2006, p. 167) also emphasise the fact that the promotion of autonomy in an HE environment takes “time and practice”, with Tschirhart & Rigler (2009, p. 82) going so far as to say that “becoming (more) autonomous is a never-ending process for all of us”.

Auerbach (2007, p. 87) reminds us that “moving towards autonomy is a bumpy ride where contradictions, uncertainty, and conflicts are obstacles to be expected and overcome” and it could be said that the AARP was one such “bumpy ride”. It cannot be claimed that the AARP was welcomed by all participants with equal relish, as an opportunity to enhance their language learning, through exercising a greater degree of autonomy; however, it does seem reasonable to claim that it did encourage greater reflection and critical thinking in all participants and, hence, for the duration of at least one of their university courses, a lesser degree of heteronomy, something that will stand them in good stead through a lifetime of learning, and, in the case of the majority of participants, a lifetime of teaching.
BIBLIOGRAPHY


Ignatieff, M. (1990, January 7). We are the artists of our own lives. The Observer.


Little, D. (1999c). Learner autonomy is more than a Western cultural concept. In S. Cotterall & D. Crabbe (Eds.), *Learner autonomy in language learning: Defining the field and effecting change* (pp. 11-18). Frankfurt: Peter Lang.


Newby, D. (2000). Authenticity (introductory article). In A.-B. Fenner & D. Newby (Eds.), *Approaches to materials design in European textbooks: Implementing principles of authenticity, learner autonomy, cultural awareness* (pp. 16-23). Graz: ECML.


Degrees of autonomy in foreign language learning

by

Carol J. Everhard-Theophilidou

A thesis submitted for the degree of

Doctor of Philosophy

to the

Department of Theoretical and Applied Linguistics

School of English

Aristotle University of Thessaloniki

VOLUME 2

January 2012
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Table of contents</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Appendices</td>
<td>ii</td>
</tr>
<tr>
<td>Appendices 1-10: Language Mastery I – awareness-raising, assignments and assessment</td>
<td>1</td>
</tr>
<tr>
<td>Appendices 11-16: AARP Post-Study Intervention</td>
<td>20</td>
</tr>
<tr>
<td>Appendices 17-20: Overview of assessment efficiency</td>
<td>34</td>
</tr>
<tr>
<td>Appendices 21-25: Assessment questionnaires</td>
<td>39</td>
</tr>
<tr>
<td>Appendices 26a-26j: Pre-Study questionnaire response analysis</td>
<td>50</td>
</tr>
<tr>
<td>Appendices 27a-27j: Main Study questionnaire response analysis</td>
<td>61</td>
</tr>
<tr>
<td>Appendices 28a-28v: Post-Study questionnaire response analysis</td>
<td>72</td>
</tr>
<tr>
<td>Bibliography</td>
<td>95</td>
</tr>
<tr>
<td>Musicography</td>
<td>96</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>No.</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APPENDICES 1-10</strong></td>
<td></td>
</tr>
<tr>
<td>LANGUAGE MASTERY I - AWARENESS-RAISING, ASSIGNMENTS AND ASSESSMENT</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>What kind of learner are you?</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Left- or right-brain dominance?</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Student Profile Card</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Learner-Teacher Contract</td>
<td></td>
</tr>
<tr>
<td>5a</td>
<td>10</td>
</tr>
<tr>
<td>1st Home Writing Assignment topic</td>
<td></td>
</tr>
<tr>
<td>5b</td>
<td>11</td>
</tr>
<tr>
<td>Song – Luka – Pre-listening exercise</td>
<td></td>
</tr>
<tr>
<td>5c</td>
<td>12</td>
</tr>
<tr>
<td>Song – Luka – While-listening exercise</td>
<td></td>
</tr>
<tr>
<td>5d</td>
<td>13</td>
</tr>
<tr>
<td>Song – Luka – Answer key</td>
<td></td>
</tr>
<tr>
<td>5e</td>
<td>14</td>
</tr>
<tr>
<td>Questionnaire: What kind of person are you?</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>1st Home Writing Assignment assessment criteria checklist</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>2nd Home Writing Assignment topic</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>2nd Home Writing Assignment assessment criteria checklist</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>Oral Assignment Topic – Suggestions</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>Oral Assignment assessment criteria checklist</td>
<td></td>
</tr>
<tr>
<td><strong>APPENDICES 11-16</strong></td>
<td></td>
</tr>
<tr>
<td>AARP POST-STUDY INTERVENTION</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>Mock writing assessment – Samples 1-5</td>
<td></td>
</tr>
<tr>
<td>12a</td>
<td>21</td>
</tr>
<tr>
<td>Assessment sheet for Mock Samples 1-3 – Homework</td>
<td></td>
</tr>
<tr>
<td>12b</td>
<td>26</td>
</tr>
<tr>
<td>Assessment sheet for Mock Samples 4-5 + Real In-class peer-assessment task</td>
<td></td>
</tr>
<tr>
<td>13a</td>
<td>27</td>
</tr>
<tr>
<td>Original ratings of Mock Samples 1-3 for comparison</td>
<td></td>
</tr>
<tr>
<td>13b</td>
<td>28</td>
</tr>
<tr>
<td>Original ratings of Mock Samples 4-5 for comparison</td>
<td></td>
</tr>
<tr>
<td>14a</td>
<td>29</td>
</tr>
<tr>
<td>Anonymous Sample of student’s ratings of Mock Samples 1-3, for comparison with I-R</td>
<td></td>
</tr>
<tr>
<td>14b</td>
<td>30</td>
</tr>
<tr>
<td>Anonymous Sample of student’s ratings of Mock Samples 4-5 + Real in-class peer-assessment for comparison with I-R</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>Assessment sheet for Mock oral presentations 1-9</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td>Anonymous Sample of student’s ratings of Mock presentations 1-9, for comparison with I-R</td>
<td></td>
</tr>
<tr>
<td><strong>APPENDICES 17-20</strong></td>
<td></td>
</tr>
<tr>
<td>OVERVIEW OF LEARNER ASSESSMENT SUCCESS</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>AARP assessment results overview</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>35</td>
</tr>
<tr>
<td>Chi-square tests for 1st Writing Assignment rating</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td>Chi-square tests for 2nd Writing Assignment rating</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>37</td>
</tr>
<tr>
<td>Chi-square tests for Oral Presentation rating</td>
<td></td>
</tr>
<tr>
<td>APPENDICES 21-25</td>
<td>39</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>ASSESSMENT QUESTIONNAIRES</td>
<td></td>
</tr>
<tr>
<td>21 Assessment questionnaire used in Pre- and Main Study – (authentic Greek version)</td>
<td>40</td>
</tr>
<tr>
<td>22 Assessment questionnaire used in Pre- and Main Study – (English translation)</td>
<td>42</td>
</tr>
<tr>
<td>23 Abridged form of Assessment questionnaire, used for analysis in the AARP, in English translation</td>
<td>44</td>
</tr>
<tr>
<td>24 Overview of Rogue Group E’s responses to Assessment questionnaire</td>
<td>45</td>
</tr>
<tr>
<td>25 Intervention Assessment Questionnaire (in English)</td>
<td>46</td>
</tr>
<tr>
<td>APPENDICES 26A-26J</td>
<td>50</td>
</tr>
<tr>
<td>PRE-STUDY QUESTIONNAIRE RESPONSE ANALYSIS</td>
<td></td>
</tr>
<tr>
<td>APPENDICES 27A-27J</td>
<td>61</td>
</tr>
<tr>
<td>MAIN STUDY QUESTIONNAIRE RESPONSE ANALYSIS</td>
<td></td>
</tr>
<tr>
<td>APPENDICES 28A-28V</td>
<td>72</td>
</tr>
<tr>
<td>POST-STUDY QUESTIONNAIRE RESPONSE ANALYSIS</td>
<td></td>
</tr>
</tbody>
</table>
APPENDICES 1-10
Language Mastery I – Awareness-raising, Assignments & Assessment
APPENDIX 1

What Kind of Learner Are You?

1. How can other people best interpret their emotions?
   a. through your facial expressions
   b. from the quality of your voice
   c. through your general body language

2. How do you manage to keep up with current events?
   a. by reading the newspaper thoroughly when you have the time
   b. by listening to the radio or watching the news
   c. by quickly reading the paper or spending just a few minutes watching the TV news

3. What sort of driver (or passenger) are you?
   a. you frequently check the rear view mirror and watch the road carefully
   b. you turn on the radio as soon as you get into the car
   c. you can’t get comfortable in the seat and continually shift position

4. How do you prefer to conduct business?
   a. by having face-to-face meetings or writing letters
   b. over the phone because it saves time
   c. by talking while you are walking, jogging or doing something else physical

5. How do you react when you’re angry?
   a. by clamping up and giving others the silent treatment
   b. by quickly letting others know when you’re angry
   c. by clenching your fists, grasping something tightly or storming off

6. How would you describe the way you dress?
   a. a neat and tidy dresser
   b. a sensible dresser
   c. a comfortable dresser

7. What do you think is the best way to discipline a child?
   a. to isolate the child by separating him/her from the group
   b. to reason with the child and discuss the situation
   c. to use acceptable forms of corporal punishment

8. How do you behave at meetings?
   a. you come prepared with notes and displays
   b. you enjoy discussing issues and hearing other points of view
   c. you would rather be somewhere else and so spend your time doodling

9. What do you like doing in your free time?
   a. watching TV or going to the cinema
   b. listening to the radio, going to a concert or playing a musical instrument
   c. engaging in physical activity of some kind

10. What do you consider to be the best way of rewarding students?
    a. writing positive comments on their work
    b. giving oral praise to the student
    c. a pat on the back, a hug, or some other appropriate physical action

From Berman (1998)
What Your Score Means

If most of your answers are A, then your modality strength is visual. In other words, you learn through seeing things and you like everything to be written down on paper. In a classroom, having notes and the use of visual aids will help you.

If most of your answers are B, your modality strength is auditory. In other words, you learn through listening. In a classroom you will want to hear the new language, and listening to music could well be helpful.

If most of your answers are C, your modality strength is kinesthetic. In other words, you learn on the move or through movement. Sitting passively in a classroom is unlikely to appeal to you but you’ll probably respond well to the use of games and role-play.

From Berman (1998)
APPENDIX 2

Complete the exercise to find your left-brain, right-brain orientation:-

Example:

<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>I prefer speaking to large</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>audiences.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I prefer speaking in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>small group situations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Box number 4 has been checked to indicate a moderate preference for speaking in small group situations

1. I remember names.    1. I remember faces.
2. I respond better to verbal instructions. 2. I respond better to demonstrated, illustrated, symbolic instructions.
3. I am intuitive. 3. I am intellectual.
4. I experiment randomly and with little restraint. 4. I experiment systematically and with control.
5. I prefer solving a problem by breaking it into parts, then approaching the problem sequentially, using logic. 5. I prefer solving a problem by looking at the whole, the configurations, then approaching the problem through patterns, using hunches.
6. I make objective judgments extrinsic to person. 6. I make subjective judgments, intrinsic to person.
7. I am fluid and spontaneous. 7. I am planned and structured.
8. I prefer established, certain information. 8. I prefer elusive, uncertain information.
9. I am a synthesizing reader. 9. I am an analytical reader.
10. I rely primarily on language in thinking and remembering. 10. I rely primarily on images in thinking and remembering.
11. I prefer talking and writing. 11. I prefer drawing and manipulating objects.
12. I get easily distracted trying to read a book in noisy or crowded places. 12. I can easily concentrate on reading a book in noisy or crowded places.
13. I prefer work and/or studies that are open-ended. 13. I prefer work and/or studies that are carefully planned.
15. I control my feelings. 15. I am more free with my feelings.
17. I am good at interpreting body language. 17. I am good at paying attention to people's exact words.
18. I frequently use metaphors and analogies. 18. I rarely use metaphors or analogies.
20. I prefer multiple-choice tests. 20. I prefer open-ended questions.

From Brown (1989)
Scoring directions.

Score each item as follows: Some of the items are scored according to the numbers at the top of each column of boxes, others are reversed. For the following items use the indicated numbers on the test page:

1
2
10
11

5
14

6
15

8
19
20

The rest of the items are reversed in their scoring. Score the following items using the numbers indicated at right.

3
4
9

7
12
13

16
17
18

5
4
3
2
1

Now total up all scores: __________

This was a test of left- and right-brain preference. A score of sixty is the mid-point. The scoring chart below indicates that a score of sixty plus or minus three is a toss-up:

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above 70</td>
<td>Quite right-brain oriented</td>
</tr>
<tr>
<td>64-70</td>
<td>Moderately right-brain oriented</td>
</tr>
<tr>
<td>57-63</td>
<td>No particular dominance on either side</td>
</tr>
<tr>
<td>50-56</td>
<td>Moderately left-brain oriented</td>
</tr>
<tr>
<td>Below 50</td>
<td>Quite left-brain oriented</td>
</tr>
</tbody>
</table>

Does the score seem to fit your own view of yourself? If not, go with your intuition, since tests like these aren't accurate for everyone. However you line up on left- and right-brain differences, capitalize on the strengths of your own orientation. Be aware of those strengths and don't let yourself rely too much on one type of functioning. If you scored close to the midpoint, that may be a rather good indication of your ability to function with both sides of the brain.

Now continue reading the chapter where you left off; it will help you interpret your score.

From Brown (1989)
APPENDIX 3

LEARNER PROFILE

Surname: 
First name: 
Father’s name: 
Student number: 
Place / country of origin: 
Interests / Hobbies: 
Ambitions: 

Learner / learner contract: Yes [   ] No [   ]
Support Group (1): Yes [   ] No [   ]
Coordinator: 
Other Group Members:

Support Group (2): Yes [   ] No [   ]
Coordinator: 
Other Group Members:

1st Oral Assignment Yes [   ] No [   ]
Title: 
Comments: 
Date: 
Own assessment: [   ] Peer assessment: [   ] Instructor's assessment: [   ]

2nd Oral Assignment Yes [   ] No [   ]
Topic: 
Comments: 
Date: 
Own assessment: [   ] Peer assessment: [   ] Instructor's assessment: [   ]

1st Class Assignment Comments:
Title: 
Punctual: Yes [   ] No [   ]
Comments: 
Own assessment: [   ] Peer assessment: [   ] Instructor's assessment: [   ]

2nd Class Assignment Comments:
Title: 
Punctual: Yes [   ] No [   ]
Comments: 
Own assessment: [   ] Peer assessment: [   ] Instructor's assessment: [   ]

3rd Class Assignment Comments:
Title: 
Punctual: Yes [   ] No [   ]
Comments: 
Own assessment: [   ] Peer assessment: [   ] Instructor's assessment: [   ]
1st Home Assignment (Self) Comments:
Title:
…………………………………….
Punctual: Yes [ ] No [ ]
…………………………………….
Voluntary / Obligatory
Own assessment: [ ] Peer assessment: [ ] Instructor's assessment: [ ]

1st Home Assignment (Peer) Comments:
Title:
…………………………………….
Punctual: Yes [ ] No [ ]
…………………………………….
Voluntary / Obligatory
Own assessment: [ ] Peer assessment: [ ] Instructor's assessment: [ ]

2nd Home Assignment (Self) Comments:
Title:
…………………………………….
Punctual: Yes [ ] No [ ]
…………………………………….
Voluntary / Obligatory
Own assessment: [ ] Peer assessment: [ ] Instructor's assessment: [ ]

2nd Home Assignment (Peer) Comments:
Title:
…………………………………….
Punctual: Yes [ ] No [ ]
…………………………………….
Voluntary / Obligatory
Own assessment: [ ] Peer assessment: [ ] Instructor's assessment: [ ]

Diagnostic Test Results 1st part [ ] 2nd part [ ] Total [ ] Level [ ]

EFL Certificates & Dates 1) ………….. 2) ……………….. 3) ……………….. 4) ……………….. 5) ……………….. 6) ………………..

Preferred Mode of Learning Questionnaire A [ ] B [ ] C [ ]

Left Brain / Right Brain Questionnaire Score [ ] = …………………..

Overall Speaking Grade Overall Writing Grade Grade for Participation Final Exam Grade Total Final Grade
Course 101

7
APPENDIX 4

LEARNER / TEACHER CONTRACT

Instructor’s name: ______________________
Student’s name: _______________________

Writing strengths:

Good ideas Fluency of expression Rich vocabulary Clarity
Correct syntax/grammar Cohesion Accurate spelling Coherence
Awareness of audience Accurate paragraphing Sufficient sophistication in syntax and vocabulary
Precision Correct use of punctuation Good at revising/editing/improving
Good at beginnings Good at endings Correct choice of words in context

Student’s remarks:
_____________________________________________________________________
_____________________________________________________________________
_____________________________________________________________________

Instructor’s remarks:
_____________________________________________________________________
_____________________________________________________________________
_____________________________________________________________________

Goals decided on:
_____________________________________________________________________
_____________________________________________________________________
_____________________________________________________________________

1st goal and route to achieving it:
Comments:  .................................................................
..............................................................................
..............................................................................

2nd goal and route to achieving it:
Comments:  .................................................................
..............................................................................
..............................................................................

3rd goal and route to achieving it:
Comments:  .................................................................
..............................................................................
..............................................................................

4th goal and route to achieving it:
Comments:  .................................................................
..............................................................................
..............................................................................

5th goal and route to achieving it:
Comments:  .................................................................
..............................................................................
..............................................................................
Speaking strengths:

Good at introducing ideas      Fluency of expression      Accuracy of expression
Rich vocabulary      Good at clarifying, re-expressing and giving examples
Correct syntax/ grammar      Clarity      Correct use of pauses and connecting devices
Awareness of audience and appropriacy of language      Accurate pronunciation
Correct choice of words in context      Coherence      Good at openings
Able to respond to questions appropriately      Use of body language, intonation etc
Sufficient degree of sophistication in language use
Good at rounding off / ending/ reaching a conclusion

Student’s remarks:

________________________________________

________________________________________

Instructor’s remarks:

________________________________________

________________________________________

Goals decided on:

________________________________________

________________________________________

1\textsuperscript{st} goal and route to achieving it:

Comments: ..........................................................
                      ..........................................................
                      ..........................................................

2\textsuperscript{nd} goal and route to achieving it:

Comments: ..........................................................
                      ..........................................................
                      ..........................................................

3\textsuperscript{rd} goal and route to achieving it:

Comments: ..........................................................
                      ..........................................................
                      ..........................................................

4\textsuperscript{th} goal and route to achieving it:

Comments: ..........................................................
                      ..........................................................
                      ..........................................................

5\textsuperscript{th} goal and route to achieving it:

Comments: ..........................................................
                      ..........................................................
                      ..........................................................

Date: ........  Student’s Signature: .............  Instructor’s Signature: .............
APPENDIX 5a

Language Mastery I – 1st Home Writing Assignment – Paragraph Writing

1) Describe the person from the song Luka as you imagine them to be

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

2) Using the information from the interview, describe the person you talked to at the first meeting of our class

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________
### APPENDIX 5b

**Song - Luka**

**Before listening - guessing words from context**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td></td>
</tr>
<tr>
<td>2)</td>
<td></td>
</tr>
<tr>
<td>3)</td>
<td></td>
</tr>
<tr>
<td>4)</td>
<td></td>
</tr>
<tr>
<td>5)</td>
<td></td>
</tr>
<tr>
<td>6)</td>
<td></td>
</tr>
<tr>
<td>7)</td>
<td></td>
</tr>
<tr>
<td>8)</td>
<td></td>
</tr>
<tr>
<td>9)</td>
<td></td>
</tr>
<tr>
<td>10)</td>
<td></td>
</tr>
<tr>
<td>11)</td>
<td></td>
</tr>
<tr>
<td>12)</td>
<td></td>
</tr>
<tr>
<td>13)</td>
<td></td>
</tr>
<tr>
<td>14)</td>
<td></td>
</tr>
<tr>
<td>15)</td>
<td></td>
</tr>
<tr>
<td>16)</td>
<td></td>
</tr>
</tbody>
</table>

**After listening:**

I would like / would not like to meet Luka because ….

__________________________________________________________

__________________________________________________________

Make some notes her about Luka in preparation for 1st Assignment

Paragraph Writing

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________

__________________________________________________________
**Luka**

My name is Luka
I live on the second (1) _________
I live (2) _________ from you
Yes, I think you’ve seen me (3) _________
If you hear something (4) _________ at night
Some kind of (5) _________, some kind of fight
Just don’t ask me what it was (3 times)
I think it’s (6) _________ I’m clumsy
I try not to talk too (7) _________
(8) _________ it’s because I’m crazy
I try not to (9) _________ too proud
They only hit until you (10) _________
After that you don’t ask why
You just don’t (11) _________ any more (3 times)

Yes, I think I’m okay
Walked into the (12) _________ again
If you ask “what’s that”, I’ll say
It’s not your (13) _________ anyway
I (14) _________ I’d like to be alone
With nothing broken, nothing (15) _________
Just don’t (16) _________ me how I am (3 times)
My name is Luka ……..
They only hit until you ……..
APPENDIX 5d

Luka - Suzanne Vegas - Answer Key

My name is Luka
I live on the second (1) floor
I live (2) upstairs from you
Yes, I think you’ve seen me (3) before
If you hear something (4) late at night
Some kind of (5) trouble, some kind of fight
Just don’t ask me what it was (3 times)
I think it’s (6) because I’m clumsy
I try not to talk too (7) loud
(9) Maybe it’s because I’m crazy
I try not to (9) act too proud
They only hit until you (10) cry
After that you don’t ask why
You just don’t (11) argue any more (3 times)

Yes, I think I’m okay
Walked into the (12) door again
If you ask “what’s that”, I’ll say
It’s not your (13) business anyway
I (14) guess I’d like to be alone
With nothing broken, nothing (15) thrown
Just don’t (16) ask me how I am (3 times)

My name is Luka ……..
They only hit until you ……..
APPENDIX 5e

PERSONALITY QUESTIONNAIRE:

What kind of person are you?

Do this questionnaire with a partner. First of all, give your own answers about yourself. Then find out if your partner agrees with the answers you have given.

How well do you know yourself? Answer each questions with:

<table>
<thead>
<tr>
<th></th>
<th>You</th>
<th>Your partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes definitely!</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Yes and no.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>No.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Definitely not!</td>
<td></td>
</tr>
</tbody>
</table>

Do you...

- like to get up early?
- work best in the mornings?
- like to spend time outdoors?
- like to stay up late?
- feel envious of other people?
- make decisions quickly?
- plan ahead?
- work hard?
- find it easy to make friends?
- like children?
- enjoy travelling?
- try to be careful with money?
- live for the moment?
- enjoy life?

*Note that you are going to use these answers in your First Home Writing Assignment, so pay careful attention to them!*

*From Swan (n.d.)*
# APPENDIX 6

**LANGUAGE MASTERY I - 1ST HOME WRITING ASSIGNMENT - ASSESSMENT CRITERIA**

<table>
<thead>
<tr>
<th>St. No. of Writer:</th>
<th>St. No. of Assessor:</th>
<th>Overall Grade Awarded:</th>
</tr>
</thead>
</table>

1) **WRITE a self-contained PARAGRAPH. (At home - Lukas)**

Use the following criteria & scale:
1=weak  2=Moderately weak  3=Average  4=Moderately strong  5=Strong

<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>The language was precise and contained few errors</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The paragraph had relevant and related content</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The paragraph was of right size and well constructed around its topic</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The paragraph flow was logical and helped the reader</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The paragraph maintained interest throughout</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

To calculate the grade:-
1. Add the five grades together.
2. Multiply the total by four.
3. Divide by ten.

Grade awarded for **Part One** of the assignment to PEER / SELF: ..... 

2) **WRITE a self-contained PARAGRAPH. (At home – Person interviewed)**

Use the following criteria & scale:
1=weak  2=Moderately weak  3=Average  4=Moderately strong  5=Strong

<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>The language was precise and contained few errors</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The paragraph had relevant and related content</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The paragraph was of right size and well constructed around its topic</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The paragraph flow was logical and helped the reader</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>The paragraph maintained interest throughout</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

To calculate the grade:-
1. Add the five grades together.
2. Multiply the total by four.
3. Divide by ten.

Grade awarded for **Part Two** of the assignment to PEER / SELF: ..... 

To calculate the **final** grade:-
1. Add the two grades awarded together.
2. Divide by two.

**Final grade awarded:** _______
Second Writing Assignment

a) Write a descriptive essay of the person you most admire or would most like to be.

___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________
___________________________________________________________

b) Write a paragraph describing your first meeting with a boyfriend / girlfriend / best friend
APPENDIX 8

2ND Home Writing Assignment

Overall average awarded: ……….

Code number of writer: …………….. Code number of assessor: ………

1) DESCRIPTIVE ESSAY

Use the following scale and criteria:

1 = Weak  2 = Moderately weak  3 = Average  4 = Moderately strong  5 = Strong

i) The essay conveyed an interesting message from beginning to end
   1  2  3  4  5

ii) All the information in the description was relevant and clear
   1  2  3  4  5

iii) Ideas were imaginative and nicely linked
    1  2  3  4  5

iv) Language was correct and used to good effect
    1  2  3  4  5

v) The text was well-constructed and well-organised
    1  2  3  4  5

Average for PEER: …… / for SELF: …… / from INSTRUCTOR: …..

2) NARRATIVE / DESCRIPTIVE PARAGRAPH

i) The text was of the correct length and was concise
   1  2  3  4  5

ii) The content was interesting and related to the subject
    1  2  3  4  5

iii) The ideas were well-organised and well-expressed
    1  2  3  4  5

iv) The language used was correct and appropriate
    1  2  3  4  5

v) The writer aroused and held the reader’s interest till the end
    1  2  3  4  5

Average for PEER: …… / for SELF: …… / from INSTRUCTOR: …..
APPENDIX 9

**Oral Assignment Topic - Suggestions**

Here are some topics suggested for presentations, but you can choose any topic you wish which falls roughly into the category of narrative / descriptive discourse. Remember that your presentation will be much more interesting if it has some audio-visual elements. You should allow for 3 whole minutes of speaking time and a total of 5 minutes for the presentation:

1. Describe the appearance and character of the person you most admire.
2. Compare the appearance and character of two people.
3. Describe the place which most impresses you.
4. Compare and contrast two places.
5. Describe an object in detail.
6. Compare and contrast two objects in detail.
7. Describe a situation.
8. Compare and contrast two situations.
9. Describe the funniest thing that ever happened to you.
10. Describe the most embarrassing or disappointing experience you ever had.
11. Describe a day when things turned out differently from what you expected.
12. Describe your most interesting journey.
13. Describe a film you have seen.
14. Compare two films on the same subject.
15. Give an evaluation of a song / songs.
16. Give an evaluation of the work of a singer, musician, artist, actor, director, writer or other person of historical importance.
17. Describe the life of a student in Thessaloniki.
18. Describe a celebration to which you were invited.
19. Describe a concert you attended.
20. Describe a sports event which you attended.
21. Describe the best meal you ever had.
22. Describe your first experience of eating non-Greek food.
23. Describe your first experience of cooking.
24. Describe an area of beauty or of ecological interest.
25. Describe the part of your home that you like best or any place you go to get away from it all.
APPENDIX 10

LANGUAGE MASTERY I - ORAL PRESENTATION  PEER / SELF

Name/No. of Presenter: _____________________________  St. No. of Assessor: ____________

Topic presented: _______________________________  Date: ____________

Use the following criteria & scale:
1=weak   2=Moderately weak   3=Average   4=Moderately strong   5=Strong

The presentation appears to be well-prepared
1  2  3  4  5
The content and materials are appropriate and interestingly presented
1  2  3  4  5
The order of the presentation is cohesive and flows without too much reliance on notes
1  2  3  4  5
The parts of the presentation are well-linked together
1  2  3  4  5
Ability to be understood (voice, delivery, pronunciation etc.)
1  2  3  4  5

To calculate the final grade:-
4. Add the five grades together.
5. Multiply the total by four.
6. Divide by ten.

Final grade awarded: _______
APPENDICES 11-16
AARP Post-Study Intervention
APPENDIX 11a

1) **Describe the person from the song “Luka” as you imagine him to be**

The song evolves around a guy called Luka. Luka is not a carefree person. On the contrary, his life is fraught with trouble. He declares that he is not memorable. Maybe this is because he is plain. However, this can be also interpreted as an effort to make himself deliberately invisible or discreet. In a nutshell, he tries to lead a humble existence so that he does not provoke others or incite any kind of fight. Furthermore, he is the embodiment of low self esteem. This is a result of his being frequently abused. Surprisingly, he has come to a point where he has accepted it and he has ceased to fight back. Being a victim of assault you see, is par for the course for him. As a consequence, he is on the defensive all the time. Suspicion begins to creep up on him. In the end all those hardships must have wrought a profound change in his appearance. He must look like someone tormented by life. Perhaps he is all black and blue. All in all, Luka is not a person you wish to be in his shoes.

2) **Describe the person you interviewed at our first meeting**

During our first meeting I found myself interviewing Asefah, a young Iranian girl who has come to our country thanks to an exchange program. Her cheerful disposition which indicated that she had not got any trouble in acquiring new acquaintances, won immediately my friendship. Asefah is 25 years old and has already tied the knot. Actually, she keeps a photograph of her wedding day in her wallet. They seem to be a well-matched couple. Her coming here reveals her love for travelling. Although she was rather timid and by no means garrulous, I gathered that she was an early bird; no wonder she works best in the morning. One thing I found peculiar was that even though she enjoys life, she does not live for the moment and consequently prefers not to make hasty decisions but plan ahead. I attributed that to the different culture mentalities. Anyway I wish her the best!

SAMPLE /
1) Describe the person from the song ‘Luka’ as you imagine him to be

Luka is a very different person from all the others. He is very sentimental and injured. He doesn’t want to

reveal anything about the things which possibly are

listening out night in his department. He feels lonely

and he doesn’t want this. He is very isolated from

the other people. Although he wants to have a commu-
nication with them. Even though Luka is definitely a trusting

and honest person because when someone pass through

some difficult circumstances he learns to keep his

mouth closed. In addition, Luka doesn’t want to give a

report for how he is and how he reacts. Generally,

Luka is a good person who learned to pass along

call the difficult situations.


2) Describe the person you interviewed at our first meeting

The person I interviewed at my first meeting was Nancy.

Nancy was very polite and honest to me. I was really

impressed when I heard her. It was the first time that

someone made me to elise it so much. We don’t have

the same interests but I like her very much. Nancy is

a person who likes getting up early because she can do

better on her studies. Also, she doesn’t like to go out so much to

grow up. She likes to spend time alone. In addition, Nancy isn’t

jealous with the other people. She wants to work very hard to achieve all her targets. The children is a

part of her life. And her life enjoy it as she can.

Also, to Nancy likes travelling to all over the

world. She is very careful with her money and their

management. Finally, she is a person who lives for the

moment.

Sample 2
APPENDIX 11c

1. DESCRIBE THE PERSON FROM THE SONG LUKA AS YOU IMAGINE IT

Judging from the lyrics I believe that Luka is a woman. She seems to be in her early thirties as experience is obvious in her words. I think she is being abused - beaten by her boyfriend. I imagine her so beautiful and so miserable at the same time. Even though she is aware of the fact that everybody knows what she is going through - especially the person who lives downstairs she doesn't want anyone to interfere. That makes her a strong and a proud personality but at the same time a helpless and lonely creature. She even tries to justify him being aggressive and bad-tempered by saying that it might be her fault. Only a person with low self-esteem would do something like that. Her heart is broken and she would prefer to be alone than with a man-like monster, she doesn't admit anything though, that's why she won't find a solution to her problem. She pretends that everything is ok and the noise in the middle of the night is just because she is so "clumsy"!

2. DESCRIBE THE PERSON YOU INTERVIEWED AT OUR FIRST MEETING

My partner is an early bird as she likes waking up early and working in the mornings. She really does her best at that part of the day, that means that she doesn't like staying up too late apart from weekends. She is a positive and cheerful girl who enjoys life, likes meeting people, making friends and handing out with them. All these characteristics make her outgoing, friendly and tolerant with people and mainly with children. However she is not impulsive as she doesn't get excited easily or live for the moment. She is also very decisive, optimistic, ambitious as she makes up her mind fast, she plans ahead and overcomes the difficulties and the last but not least hard working because she wouldn't have succeeded anything if she wasn't one!

SAMPLE 3
APPENDIX 11d

1) Describe the person from the song ‘Luka’ as you imagine him to be

Luka is someone who has moved to his new flat recently and has fallen in love with a girl that noticed. He has problems with his adaptation there as he is a bit awkward. Luka has understood that people are selfish and envious when they see someone successful or someone better than them. He tries to be as quiet as he can because he doesn’t want other people to know about his life. However, sometimes he is really in a blue mood and feels very lonely.

SAMPLE

2) Describe the person you interviewed at our first meeting

From the interview of Kiki, I realised that she is not an early bird as she prefers to get up late and she’s not efficient in the mornings. She likes spending time outdoors, going for a walk in the countryside and hates staying up late. Kiki sometimes feels jealous of other people and she can understand this feeling from their behaviour. She is a quick thinker and plans ahead for being a good English teacher too. Taking care of children is something she really likes. Kiki is a hard worker and quite prudent especially with money. Travelling around the world is a pleasurable activity for her. She is a very sociable and ambitious person who enjoys every single moment.
1) Describe the person from the song "Luka" as you imagine him to be.

Luka seems to be a troubled figure experiencing problems in his life. His own description of himself indicates that he is a teenager going through changes that make him very judgemental about himself. So despite his need for human contact, Luka is a low-profile young man who is unwilling to share his problems with others, probably because he is too proud, as he implies, or because he is too much of an introvert. As a result, he is alone and a loner. Additionally, he seems to be passively accepting whatever goes on in his life without arguing.

SAMPLE 6

2) Describe the person you interviewed at our first meeting.

The person I interviewed at our first meeting is a low-profile young man who although might seem shy or aloof at first, he proves to be very friendly and social, as there is always a smile on his face. He is also quite a diplomat and very polite. That can also indicate that he is intelligent. He likes sports, especially football, but he is not a fanatic. Additionally, he has an absolutely great taste in books.
## APPENDIX 12a

Writing Intervention exercise – homework, with Samples 1, 2 & 3

<table>
<thead>
<tr>
<th>Your surname (in Greek)</th>
<th>Your first name (in Greek)</th>
<th>Your group (circle as appropriate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUKA</td>
<td></td>
<td>a / b</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FIRST SAMPLE</th>
<th>SECOND SAMPLE</th>
<th>THIRD SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Criterion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Criterion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Criterion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th Criterion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th Criterion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comments about paragraph</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paragraph about partner</th>
<th>FIRST SAMPLE</th>
<th>SECOND SAMPLE</th>
<th>THIRD SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comments about paragraph</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Final assignment grade</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
APPENDIX 12b

Writing Intervention exercise – in-class, with Samples 4 & 5, followed by real peer-assessment

<table>
<thead>
<tr>
<th>LUKA</th>
<th>FOURTH SAMPLE</th>
<th>FIFTH SAMPLE</th>
<th>PEER ASSESSMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ST CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2ND CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3RD CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4TH CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5TH CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVERALL GRADE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMMENTS ABOUT PARAGRAPH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PARAGRAPH ABOUT PARTNER</td>
<td>FOURTH SAMPLE</td>
<td>FIFTH SAMPLE</td>
<td>PEER ASSESSMENT</td>
</tr>
<tr>
<td>1ST CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2ND CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3RD CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4TH CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5TH CRITERION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVERALL GRADE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMMENTS ABOUT PARAGRAPH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FINAL ASSIGNMENT GRADE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 13a

Writing Intervention exercise - Previous assessment of Writing Samples 1, 2 & 3 given to students for comparison after theirs.

| LUKA | FIRST SAMPLE | | | | | | SECOND SAMPLE | | | | | | THIRD SAMPLE | | | |
|------|--------------|---|---|---|---|---|---|------------|---|---|---|---|---|---|---|---|---|---|---|---|
|      | 23 | INSTR. | PEER | SELF | 34 | INSTR. | PEER | SELF | 22 | INSTR. | PEER | SELF |
| 1ST CRITERION | 4 | 5 | 5 | 3 | 4 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2ND CRITERION | 5 | 4 | 5 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3RD CRITERION | 5 | 4 | 5 | 3 | 4 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| 4TH CRITERION | 5 | 4 | 5 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 |
| 5TH CRITERION | 5 | 4 | 5 | 3 | 4 | 3 | 5 | 5 | 4 | 4 | 5 | 5 |
| OVERALL GRADE | 9.6 | 8.4 | 10 | 6.8 | 8 | 6.8 | 10 | 10 | 9.6 |
| COMMENTS ABOUT PARAGRAPH | | | | | | | | | | | | |

| PARAGRAPH ABOUT PARTNER | FIRST SAMPLE | | | | | | SECOND SAMPLE | | | | | | THIRD SAMPLE | | | |
|-------------------------|--------------|---|---|---|---|---|---|------------|---|---|---|---|---|---|---|---|---|---|---|---|
|                         | 23 | INSTR. | PEER | SELF | 34 | INSTR. | PEER | SELF | 22 | INSTR. | PEER | SELF |
| 1ST CRITERION | 5 | 5 | 5 | 3 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 5 |
| 2ND CRITERION | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3RD CRITERION | 4 | 4 | 5 | 3 | 4 | 4 | 5 | 4 | 5 | 5 | 5 | 5 |
| 4TH CRITERION | 5 | 4 | 5 | 4 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 5 |
| 5TH CRITERION | 5 | 4 | 5 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| OVERALL GRADE | 9.2 | 8.4 | 9.6 | 6.8 | 6.8 | 6.8 | 9.2 | 8.8 | 10 | 9.2 | 8.8 | 10 |
| COMMENTS ABOUT PARAGRAPH | | | | | | | | | | | | |

| FINAL ASSIGNMENT GRADE | 9.4 | 8.4 | 9.8 | 6.8 | 7.4 | 6.8 | 9.6 | 9.4 | 9.8 | 9.6 | 9.4 | 9.8 |
### APPENDIX 13b

Writing Intervention exercise - Previous assessment of Writing Samples 4 & 5 given to students for comparison after theirs.

<table>
<thead>
<tr>
<th>LUKA</th>
<th>FOURTH SAMPLE</th>
<th></th>
<th>FIFTH SAMPLE</th>
<th></th>
<th>PEER ASSESSMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>32</td>
<td>INSTR.</td>
<td>PEER</td>
<td>SELF</td>
<td>No.</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; CRITERION</td>
<td>3 4 4</td>
<td>4 4 4</td>
<td>5 4 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; CRITERION</td>
<td>4 4 3</td>
<td>3 4 3</td>
<td>5 4 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; CRITERION</td>
<td>4 3 3</td>
<td>3 4 4</td>
<td>5 4 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; CRITERION</td>
<td>4 3 3</td>
<td>3 4 4</td>
<td>5 4 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt; CRITERION</td>
<td>4 3 2</td>
<td>3 4 4</td>
<td>5 4 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OVERALL GRADE</strong></td>
<td>7.6 6.8 6</td>
<td>10 6.8 8.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**COMMENTS ABOUT PARAGRAPH**

<table>
<thead>
<tr>
<th>PARAGRAPH ABOUT PARTNER</th>
<th>FOURTH SAMPLE</th>
<th></th>
<th>FIFTH SAMPLE</th>
<th></th>
<th>PEER ASSESSMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>INSTR.</td>
<td>PEER</td>
<td>SELF</td>
<td>No.</td>
<td></td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; CRITERION</td>
<td>4 4 4</td>
<td>3 3 4</td>
<td>3 3 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; CRITERION</td>
<td>4 5 3</td>
<td>3 2 4</td>
<td>3 2 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; CRITERION</td>
<td>4 4 4</td>
<td>3 2 4</td>
<td>3 2 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; CRITERION</td>
<td>4 3 4</td>
<td>3 3 4</td>
<td>3 3 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt; CRITERION</td>
<td>4 4 4</td>
<td>3 3 4</td>
<td>3 3 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OVERALL GRADE</strong></td>
<td>8 8.4 7.2</td>
<td>6 4.8 6.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FINAL GRADE**

|              | 7.8 7.6 6.6 | 8 5.8 7.6 |            |           |
### APPENDIX 14a

**Writing Intervention exercise - ASSESSOR: ANONYMOUS***

<table>
<thead>
<tr>
<th>LUKA</th>
<th>FIRST SAMPLE</th>
<th>INSTR.</th>
<th>PEER</th>
<th>SELF</th>
<th>23</th>
<th>34</th>
<th>INSTR.</th>
<th>PEER</th>
<th>SELF</th>
<th>22</th>
<th>INSTR.</th>
<th>PEER</th>
<th>SELF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st CRITERION</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2nd CRITERION</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>3rd CRITERION</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>4th CRITERION</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5th CRITERION</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>OVERALL GRADE</td>
<td>10</td>
<td>9.6</td>
<td>8.4</td>
<td>10</td>
<td>7.2</td>
<td>6.8</td>
<td>8</td>
<td>6.8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9.6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARAGRAPH ABOUT PARTNER</th>
<th>FIRST SAMPLE</th>
<th>INSTR.</th>
<th>PEER</th>
<th>SELF</th>
<th>23</th>
<th>34</th>
<th>INSTR.</th>
<th>PEER</th>
<th>SELF</th>
<th>22</th>
<th>INSTR.</th>
<th>PEER</th>
<th>SELF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st CRITERION</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2nd CRITERION</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>3rd CRITERION</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4th CRITERION</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5th CRITERION</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>OVERALL GRADE</td>
<td>8</td>
<td>9.2</td>
<td>8.4</td>
<td>9.6</td>
<td>6.8</td>
<td>6.8</td>
<td>6.8</td>
<td>6.8</td>
<td>10</td>
<td>9.2</td>
<td>8.8</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

*The student’s name has been removed to preserve anonymity. The student’s ratings are in bold. Colour highlighting shows where P-A and I-A scoring of criteria coincide. The Instructor, Peer and Self-ratings given are the authentic ratings from the original assessment conducted in a previous year.
<table>
<thead>
<tr>
<th><strong>LUKA</strong></th>
<th><strong>FOURTH SAMPLE</strong></th>
<th><strong>FIFTH SAMPLE</strong></th>
<th><strong>ACTUAL PEER ASSESSMENT No. 20</strong></th>
<th><strong>Page 2</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INSTR. PEER SELF</td>
<td>INSTR. PEER SELF</td>
<td>INSTR. PEER SELF</td>
<td>INSTR. PEER SELF</td>
</tr>
<tr>
<td><strong>1ST CRITERION</strong></td>
<td>4 3 4 4</td>
<td>5 5</td>
<td>4 4 4</td>
<td>3 4 2</td>
</tr>
<tr>
<td><strong>2ND CRITERION</strong></td>
<td>4 4 4 3</td>
<td>5 5</td>
<td>4 5 5</td>
<td>2 5 3</td>
</tr>
<tr>
<td><strong>3RD CRITERION</strong></td>
<td>4 4 3 3</td>
<td>5 5</td>
<td>4 4 5</td>
<td>2 5 4</td>
</tr>
<tr>
<td><strong>4TH CRITERION</strong></td>
<td>4 4 3 3</td>
<td>5 5</td>
<td>2 4 4</td>
<td>3 4 3</td>
</tr>
<tr>
<td><strong>5TH CRITERION</strong></td>
<td>4 4 3 2</td>
<td>5 5</td>
<td>3 4 4</td>
<td>4 4 3</td>
</tr>
<tr>
<td><strong>OVERALL GRADE</strong></td>
<td>8 7.6 6.8 6</td>
<td>10 10</td>
<td>6.8 8.4</td>
<td>8.8 5.6</td>
</tr>
<tr>
<td><strong>PARAGRAPHS ABOUT PARTNER</strong></td>
<td><strong>FOURTH SAMPLE</strong></td>
<td><strong>FIFTH SAMPLE</strong></td>
<td><strong>ACTUAL PEER ASSESSMENT</strong></td>
<td><strong>Page 2</strong></td>
</tr>
<tr>
<td></td>
<td>INSTR. PEER SELF</td>
<td>INSTR. PEER SELF</td>
<td>INSTR. PEER SELF</td>
<td>INSTR. PEER SELF</td>
</tr>
<tr>
<td><strong>1ST CRITERION</strong></td>
<td>4 4 4 4</td>
<td>3 3</td>
<td>3 4 4</td>
<td>3 4 4</td>
</tr>
<tr>
<td><strong>2ND CRITERION</strong></td>
<td>4 4 5 3</td>
<td>3 3</td>
<td>2 4 5</td>
<td>4 5 4</td>
</tr>
<tr>
<td><strong>3RD CRITERION</strong></td>
<td>4 4 4 4</td>
<td>3 3</td>
<td>2 2 5</td>
<td>3 5 5</td>
</tr>
<tr>
<td><strong>4TH CRITERION</strong></td>
<td>4 4 3 3</td>
<td>2 3 2 4</td>
<td>5 4 5 4</td>
<td></td>
</tr>
<tr>
<td><strong>5TH CRITERION</strong></td>
<td>4 4 4 4</td>
<td>3 3 3 3</td>
<td>5 4 5 4</td>
<td></td>
</tr>
<tr>
<td><strong>OVERALL GRADE</strong></td>
<td>8 8 8.4 7.2</td>
<td>5.6 6</td>
<td>4.8 6.8</td>
<td>9.6 7.2</td>
</tr>
<tr>
<td><strong>FINAL ASSIGNMENT GRADE</strong></td>
<td>8 7.8 7.6 6.6</td>
<td>7.8 8</td>
<td>5.8 7.6</td>
<td>9.2 6.4</td>
</tr>
</tbody>
</table>

*The student’s name has been removed to preserve anonymity. The student’s ratings are in bold. Colour highlighting shows where P-A and I-A scoring of criteria coincide. The Instructor, Peer and Self-ratings given are the authentic ratings from the original assessment conducted in a previous year, with the exception of Actual Peer Assessment, where Actual Peer and Peer are one and the same.
### APPENDIX 15

Oral Intervention exercise – assessment record sheet (1) for Mock Presentations

<table>
<thead>
<tr>
<th>MOC PRESENTATIONS</th>
<th>FIRST PRESENTER</th>
<th>SECOND PRESENTER</th>
<th>THIRD PRESENTER</th>
<th>FOURTH PRESENTER</th>
<th>FIFTH PRESENTER</th>
<th>SIXTH PRESENTER</th>
<th>SEVENTH PRESENTER</th>
<th>EIGHTH PRESENTER</th>
<th>NINTH PRESENTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ST CRITERION</td>
<td>MARINA</td>
<td>VENETIA</td>
<td>OLGA</td>
<td>PANAGIOTA</td>
<td>ELENI</td>
<td>KATERINA</td>
<td>MAJA</td>
<td>MARIA</td>
<td>MENIA</td>
</tr>
<tr>
<td>2ND CRITERION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3RD CRITERION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4TH CRITERION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5TH CRITERION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OVERALL GRADE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMMENTS ABOUT PRESENTATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### APPENDIX 16

**Oral Intervention exercise – assessment record sheet**

**Student Assessor:** ANONYMOUS*

<table>
<thead>
<tr>
<th>MOCK PRESENTATIONS</th>
<th>FIRST PRESENTER</th>
<th>SECOND PRESENTER</th>
<th>THIRD PRESENTER</th>
<th>FOURTH PRESENTER</th>
<th>FIFTH PRESENTER</th>
<th>SIXTH PRESENTER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MARINA INSTR.</td>
<td>VENETIA INSTR.</td>
<td>OLGA INSTR.</td>
<td>PANAGIOTA INSTR.</td>
<td>ELENI INSTR.</td>
<td>KATERINA INSTR.</td>
</tr>
<tr>
<td>1&lt;sup&gt;ST&lt;/sup&gt; CRITERION</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2&lt;sup&gt;ND&lt;/sup&gt; CRITERION</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>3&lt;sup&gt;RD&lt;/sup&gt; CRITERION</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4&lt;sup&gt;TH&lt;/sup&gt; CRITERION</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5&lt;sup&gt;TH&lt;/sup&gt; CRITERION</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>OVERALL GRADE</td>
<td>6</td>
<td>6.8</td>
<td>9.2</td>
<td>8.8</td>
<td>9.2</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MOCK PRESENTATIONS</th>
<th>SEVENTH PRESENTER</th>
<th>EIGHTH PRESENTER</th>
<th>NINTH PRESENTER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAJA INSTR. MARIA</td>
<td>INSTR. MENIA</td>
<td>INSTR.</td>
</tr>
<tr>
<td>1&lt;sup&gt;ST&lt;/sup&gt; CRITERION</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2&lt;sup&gt;ND&lt;/sup&gt; CRITERION</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3&lt;sup&gt;RD&lt;/sup&gt; CRITERION</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4&lt;sup&gt;TH&lt;/sup&gt; CRITERION</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5&lt;sup&gt;TH&lt;/sup&gt; CRITERION</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>OVERALL GRADE</td>
<td>10</td>
<td>10</td>
<td>7.2</td>
</tr>
</tbody>
</table>

*The student’s name has been removed to preserve anonymity. The student’s ratings are in bold. Colour highlighting shows where P-A and I-A scoring of criteria coincide.
APPENDICES 17-20
Overview of assessment proficiency
## APPENDIX 17

### AARP Assessment Overview

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-6 Group A</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-6 Group B</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-7 Group C</td>
<td>sig. ( p = 0.002 ) A&gt;C=B</td>
<td>0.62</td>
<td>0.61</td>
<td>n.s.</td>
<td>0.48</td>
<td>n.s.</td>
<td>0.45</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-7 Group D</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-8 Group E</td>
<td>sig. ( p = 0.006 ) A&gt;C only</td>
<td>0.51</td>
<td>0.50</td>
<td>n.s.</td>
<td>0.46</td>
<td>0.42</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-8 Group F</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-9 Group G</td>
<td>sig. ( p = 0.005 ) A&gt;B only</td>
<td>sig. ( p = 0.048 ) A&gt;B only</td>
<td>0.47</td>
<td>n.s.</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008-9 Group H</td>
<td>n.s.</td>
<td>sig. ( p &lt;0.017 ) A&gt;B only</td>
<td>0.66</td>
<td>n.s.</td>
<td>0.52</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-10* Group I</td>
<td>sig. ( p &lt;0.001 ) A&gt;B&gt;C</td>
<td>sig. ( p = 0.021 ) A&gt;C only</td>
<td>0.65</td>
<td>n.s.</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009-10* Group J</td>
<td>n.s.</td>
<td>0.66</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

n.s. = non-significant  
 sig. = significant  
 ANOVA - A = Self  
 B = Peer  
 C = Instructor  
 PEARSON - S-I = Self-Instructor  
 P-I = Peer-Instructor  
 P-S = Peer-Self  
 2009-2010* = The results and values offered here are based only on participants in the Post-Study Intervention exercises. Non-participants have been excluded.
## APPENDIX 18

Chi-square tests for 1st Home Writing Assignment

<table>
<thead>
<tr>
<th>Year</th>
<th>I</th>
<th>S</th>
<th>Total</th>
<th>Chi-sq</th>
<th>E</th>
<th>I</th>
<th>S</th>
<th>Total</th>
<th>Chi-sq</th>
<th>E</th>
<th>I</th>
<th>S</th>
<th>Total</th>
<th>Chi-sq</th>
<th>E</th>
<th>I</th>
<th>S</th>
<th>Total</th>
<th>Chi-sq</th>
<th>E</th>
<th>I</th>
<th>S</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/06</td>
<td>27</td>
<td>14</td>
<td>41</td>
<td>13,562</td>
<td>27,52</td>
<td>20</td>
<td>13</td>
<td>42</td>
<td>1,126</td>
<td>15,78</td>
<td>22</td>
<td>24</td>
<td>44</td>
<td>0,63</td>
<td>15,78</td>
<td>22</td>
<td>24</td>
<td>44</td>
<td>0,63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006/07</td>
<td>12,49</td>
<td>25,51</td>
<td>38</td>
<td>0,498</td>
<td>0,244</td>
<td>12</td>
<td>34</td>
<td>46</td>
<td>0,385</td>
<td>0,215</td>
<td>12</td>
<td>28</td>
<td>40</td>
<td>0,244</td>
<td>0,385</td>
<td>0,215</td>
<td>0,244</td>
<td>0,385</td>
<td>0,215</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007/08</td>
<td>15,12</td>
<td>30,88</td>
<td>46</td>
<td>0,645</td>
<td>0,316</td>
<td>15</td>
<td>31</td>
<td>46</td>
<td>0,505</td>
<td>0,283</td>
<td>15</td>
<td>31</td>
<td>46</td>
<td>0,283</td>
<td>0,505</td>
<td>0,283</td>
<td>0,283</td>
<td>0,505</td>
<td>0,283</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008/09</td>
<td>15,12</td>
<td>30,88</td>
<td>46</td>
<td>0,001</td>
<td>0,001</td>
<td>8</td>
<td>40</td>
<td>48</td>
<td>0,137</td>
<td>0,077</td>
<td>8</td>
<td>40</td>
<td>48</td>
<td>0,077</td>
<td>0,137</td>
<td>0,077</td>
<td>0,077</td>
<td>0,137</td>
<td>0,077</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009/10</td>
<td>15,78</td>
<td>32,22</td>
<td>48</td>
<td>3,836</td>
<td>1,879</td>
<td>15</td>
<td>31</td>
<td>46</td>
<td>0,602</td>
<td>0,337</td>
<td>15</td>
<td>31</td>
<td>46</td>
<td>0,337</td>
<td>0,602</td>
<td>0,337</td>
<td>0,337</td>
<td>0,602</td>
<td>0,337</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>147</td>
<td>219</td>
<td></td>
<td></td>
<td>104</td>
<td>117</td>
<td>221</td>
<td></td>
<td></td>
<td>80</td>
<td>143</td>
<td>223</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sum Chi-Sq = 26.623; DF = 4; P<0.001
Sum Chi-Sq = 20.820; DF = 4; P<0.001
Sum Chi-Sq = 4.296; DF = 4; P=0.367

- **O** = Observed counts
- **E** = Expected counts
- **Ch-sq** = Chi-Square contributions

I = Instructor  S = Self  P = Peer
# APPENDIX 19

Chi-square tests for 2nd Home Writing Assignment

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>S</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/2006</td>
<td>O</td>
<td>15</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>12.2</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.643</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>10.78</td>
<td>27.22</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>3.1</td>
<td>1.228</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>11</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>11.35</td>
<td>28.65</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.011</td>
<td>0.004</td>
</tr>
<tr>
<td>2006/2007</td>
<td>O</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>12.77</td>
<td>32.23</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.046</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>18</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>13.9</td>
<td>35.1</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>1.208</td>
<td>0.478</td>
</tr>
<tr>
<td>2007/2008</td>
<td>O</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>20.41</td>
<td>21.59</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>18.47</td>
<td>19.53</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.651</td>
<td>0.615</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>10.91</td>
<td>26.09</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.335</td>
<td>0.14</td>
</tr>
<tr>
<td>2008/2009</td>
<td>O</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>21.87</td>
<td>23.13</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>11</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>13.9</td>
<td>35.1</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>1.165</td>
<td>1.102</td>
</tr>
<tr>
<td>2009/2010</td>
<td>O</td>
<td>28</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>22.84</td>
<td>24.16</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.054</td>
<td>0.022</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>P</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/2006</td>
<td>O</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>20.41</td>
<td>21.59</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>2006/2007</td>
<td>O</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>18.47</td>
<td>19.53</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.651</td>
<td>0.615</td>
</tr>
<tr>
<td>2007/2008</td>
<td>O</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>21.87</td>
<td>23.13</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2008/2009</td>
<td>O</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>21.87</td>
<td>23.13</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2009/2010</td>
<td>O</td>
<td>28</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>22.84</td>
<td>24.16</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.054</td>
<td>0.022</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>S</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/2006</td>
<td>O</td>
<td>16</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>12.98</td>
<td>31.02</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.794</td>
<td>0.295</td>
</tr>
<tr>
<td>2006/2007</td>
<td>O</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>10.91</td>
<td>26.09</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.335</td>
<td>0.14</td>
</tr>
<tr>
<td>2007/2008</td>
<td>O</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>12.39</td>
<td>29.61</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.012</td>
<td>0.005</td>
</tr>
<tr>
<td>2008/2009</td>
<td>O</td>
<td>14</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>13.86</td>
<td>33.14</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>2009/2010</td>
<td>O</td>
<td>13</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>13.86</td>
<td>33.14</td>
</tr>
<tr>
<td></td>
<td>Chi-sq</td>
<td>0.054</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Total 61 154 215 Total 104 110 214 Total 64 153 217

Sum Chi-Sq = 6.991; DF = 4; P = 0.136

Sum Chi-Sq = 3.741; DF = 4; P = 0.442

Sum Chi-Sq = 1.569; DF = 4; P = 0.814

O = Observed counts
E = Expected counts
Chi-sq = Chi-Square contributions

I = Instructor  S = Self  P = Peer
Chi-square tests for Oral Assignment

<table>
<thead>
<tr>
<th>Year/Year</th>
<th>I</th>
<th>S</th>
<th>Total</th>
<th>I</th>
<th>P</th>
<th>Total</th>
<th>p</th>
<th>S</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005/2006</td>
<td>O 21</td>
<td>26</td>
<td>47</td>
<td>O 16</td>
<td>33</td>
<td>49</td>
<td>O 29</td>
<td>22</td>
<td>51</td>
</tr>
<tr>
<td>E 15.02</td>
<td>31.98</td>
<td></td>
<td></td>
<td>E 18.94</td>
<td>30.06</td>
<td></td>
<td>E 20.23</td>
<td>30.77</td>
<td></td>
</tr>
<tr>
<td>Chi-sq 2.378</td>
<td>1.117</td>
<td></td>
<td></td>
<td>Chi-sq 0.457</td>
<td>0.288</td>
<td></td>
<td>Chi-sq 3.8</td>
<td>2.499</td>
<td></td>
</tr>
<tr>
<td>2006/2007</td>
<td>O 10</td>
<td>27</td>
<td>37</td>
<td>O 17</td>
<td>23</td>
<td>40</td>
<td>O 15</td>
<td>26</td>
<td>41</td>
</tr>
<tr>
<td>E 11.83</td>
<td>25.17</td>
<td></td>
<td></td>
<td>E 15.46</td>
<td>24.54</td>
<td></td>
<td>E 16.26</td>
<td>24.74</td>
<td></td>
</tr>
<tr>
<td>Chi-sq 0.282</td>
<td>0.133</td>
<td></td>
<td></td>
<td>Chi-sq 0.153</td>
<td>0.096</td>
<td></td>
<td>Chi-sq 0.098</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>2007/2008</td>
<td>O 10</td>
<td>40</td>
<td>50</td>
<td>O 17</td>
<td>36</td>
<td>53</td>
<td>O 18</td>
<td>35</td>
<td>53</td>
</tr>
<tr>
<td>E 15.98</td>
<td>34.02</td>
<td></td>
<td></td>
<td>E 20.49</td>
<td>32.51</td>
<td></td>
<td>E 21.02</td>
<td>31.98</td>
<td></td>
</tr>
<tr>
<td>Chi-sq 2.239</td>
<td>1.052</td>
<td></td>
<td></td>
<td>Chi-sq 0.594</td>
<td>0.374</td>
<td></td>
<td>Chi-sq 0.436</td>
<td>0.266</td>
<td></td>
</tr>
<tr>
<td>Chi-sq 0.185</td>
<td>0.087</td>
<td></td>
<td></td>
<td>Chi-sq 0.322</td>
<td>0.203</td>
<td></td>
<td>Chi-sq 0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2009/2010</td>
<td>O 14</td>
<td>29</td>
<td>43</td>
<td>O 21</td>
<td>27</td>
<td>48</td>
<td>O 15</td>
<td>34</td>
<td>49</td>
</tr>
<tr>
<td>E 13.74</td>
<td>29.26</td>
<td></td>
<td></td>
<td>E 18.55</td>
<td>29.45</td>
<td></td>
<td>E 19.44</td>
<td>29.56</td>
<td></td>
</tr>
<tr>
<td>Chi-sq 0.005</td>
<td>0.002</td>
<td></td>
<td></td>
<td>Chi-sq 0.322</td>
<td>0.203</td>
<td></td>
<td>Chi-sq 1.013</td>
<td>0.666</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>149</td>
<td>219</td>
<td>Total</td>
<td>92</td>
<td>146</td>
<td>238</td>
<td>Total</td>
<td>96</td>
</tr>
</tbody>
</table>

Sum Chi-Sq = 7.479; DF = 4; P = 0.113
Sum Chi-Sq = 3.012; DF = 4; P = 0.556
Sum Chi-Sq = 8.863; DF = 4; P = 0.065

O = Observed counts
E = Expected counts
Chi-sq = Chi-Square contributions

I = Instructor  S = Self  P = Peer
APPENDICES 21-25
Assessment Questionnaires
APPENDIX 21

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ ΓΙΑ ΤΙΣ ΦΟΙΤΗΤΡΙΕΣ/ΤΟΥΣ ΦΟΙΤΗΤΕΣ

Α˚ ΜΕΡΟΣ

ΟΝΟΜΑ: (υποχρεωτικά) __________________________________________________________

ΦΥΛΟ:  Άνδρας ☐  Γυναίκα ☐

ΗΛΙΚΙΑ:  18-20 ☐  21-25 ☐  26-30 ☐  31-35 ☐  36-40 ☐

ΤΟΠΟΣ ΚΑΤΑΓΩΓΗΣ

ΧΩΡΑ: ___________________________________________ ΝΟΜΟΣ: __________________________

ΠΟΛΗ: ___________________________________________ ΚΩΜΟΠΟΛΗ: ____________________

Σημειώστε με σταυρό (+) ή με ✓ την περίπτωση που σας ταιριάζει

ΕΚΠΑΙΣΕΥΣΗ

Απολυτήριο λυκείου ☐  Πτυχίο άλλης σχολής ☐  Μεταπτυχιακός τίτλος σπουδών ☐

Άλλο: ________________________________________________________________

Έμαθα αγγλικά ☐  από τους γονείς μου ☐  με ιδιαίτερα μαθήματα ☐  στο σχολείο ☐

σε φροντιστήριο ☐  Άλλο: __________________________________________________________

Άλλες γλώσσες που γνωρίζω, εκτός από ελληνικά και αγγλικά

Γαλλικά ☐  Γερμανικά ☐  Ισπανικά ☐  Ιταλικά ☐  Άλλες: __________________________________________________________

Β˚ ΜΕΡΟΣ

1 = καθόλου  2 = ελάχιστα  3 = σχετικά  4 = αρκετά  5 = πάρα πολύ

Σημειώστε με κύκλο το βαθμό που σας ταιριάζει.

<table>
<thead>
<tr>
<th>Ερώτηση</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Γνώριζα τον κατάλογο κριτηρίων αξιολόγησης (rubrics/rating scales)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Είναι ενδιαφέρον τρόπος αξιολόγησης</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Είναι εύκολο να καταλάβω αυτόν τον τρόπο αξιολόγησης</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Καταλαβαίνω γιατί γίνεται μ’ αυτόν τον τρόπο η αξιολόγηση.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ήταν εύκολο να αξιολογήσω τον εαυτό μου</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ήταν εύκολο να αξιολογήσω την εργασία των συμφοιτητριών/τών μου.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Προτιμώ την παραδοσιακή αξιολόγηση.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

40
Ο κατάλογος κριτηρίων αξιολόγησης (rating scales/rubrics, checklists) με βοήθεια

<table>
<thead>
<tr>
<th>Κριτήριο</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>να συμμετέχω ενεργώς στο μάθημα.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να καταλαβαίνω τους στόχους του κάθε μαθήματος.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να καταλαβαίνω τι ακριβώς εξετάζοταν.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να προσέξω σημεία που αλλιώς δε θα πρόσεχα.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να αναπτύξω κριτική σκέψη για τις εργασίες μου.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να αναπτύξω κριτική σκέψη για τις εργασίες των συμφοιτητών/ τριών.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να αξιολογώ αντικειμενικά τον εαυτό μου.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να αξιολογώ αντικειμενικά τους συμφοιτητές μους/τις συμφοιτητρίες μου.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να παρακολουθώ/να διαπιστώνω την πρόοδό μου.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να γνωρίζω σε τι ακριβώς αξιολογούμαι.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να οργανώω καλύτερα τη δουλειά μου.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Θα ήθελα</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να υπήρχαν περισσότερα κριτήρια.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να κάνω περισσότερες εργασίες.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να αξιολογήσω περισσότερες εργασίες.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>να αξιολογούμαι μ’ αυτόν τον τρόπο σε όλα τα μαθήματα.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Γ’ ΜΕΡΟΣ

Παρακαλούμε γράψτε στον παρακάτω χώρο τα σχόλια σας για την τεχνική αξιολόγησης που γνωρίσατε, προσπαθώντας να δώσετε απαντήσεις στις παρακάτω ερωτήσεις αλλά και δίνοντας κάποιες ιδέες που θα μπορούσαν να βοηθήσουν την έρευνα:
1. Ποια ήταν τα δυνατά σημεία της τεχνικής;
2. Ποια ήταν τα προβλήματα που αντιμετωπίσατε;
3. Τι θα προτείνατε για τη βελτίωση της τεχνικής;

______________________________________________________________________________________
______________________________________________________________________________________
______________________________________________________________________________________
______________________________________________________________________________________
______________________________________________________________________________________
______________________________________________________________________________________
______________________________________________________________________________________
______________________________________________________________________________________
______________________________________________________________________________________

41
**APPENDIX 22**

**QUESTIONNAIRE FOR STUDENTS**

**PART ‘A’**

<table>
<thead>
<tr>
<th>NAME: (compulsory)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER: MALE</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER: FEMALE</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-20</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-25</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26-30</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-35</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-40</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**BIRTHPLACE**

<table>
<thead>
<tr>
<th>COUNTRY:</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>REGION:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CITY:</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TOWN:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Use (+) or ✓ in the option that corresponds to your situation*

**EDUCATION**

<table>
<thead>
<tr>
<th>High School Graduate</th>
<th>☐</th>
<th>Degree from different department</th>
<th>☐</th>
<th>Postgraduate degree</th>
<th>☐</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**How you learned English**

<table>
<thead>
<tr>
<th>Self-taught</th>
<th>☐</th>
<th>Taught by parents</th>
<th>☐</th>
<th>Private tutoring</th>
<th>☐</th>
<th>At School</th>
<th>☐</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tutorial classes</td>
<td>☐</td>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Other Languages you know, apart from Greek and English**

<table>
<thead>
<tr>
<th>French</th>
<th>☐</th>
<th>German</th>
<th>☐</th>
<th>Spanish</th>
<th>☐</th>
<th>Italian</th>
<th>☐</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PART ‘B’**

**FEEDBACK QUESTIONNAIRE (Part ‘B’)**

Circle the numbers that represent how you feel about aspects of the assessment.

1 = Completely disagree  2 = Tend to disagree  3 = Undecided  4 = Tend to agree  5 = Completely agree

1. I was already familiar with the criteria checklists.

   1 2 3 4 5

2. I found it an interesting form of assessment.

   1 2 3 4 5

3. I found it easy to understand this method of assessment.

   1 2 3 4 5

4. I understood why this form of assessment was being used.

   1 2 3 4 5

5. I found it easy to assess myself (self-assessment).

   1 2 3 4 5

6. I found it easy to assess the assignments of my classmates (peer-assessment).

   1 2 3 4 5

7. I prefer traditional forms of assessment.

   1 2 3 4 5
The checklists and criteria helped me to:

8. participate actively in the lessons.
   1 2 3 4 5
9. understand the aims of each lesson.
   1 2 3 4 5
10. understand what exactly was being assessed.
    1 2 3 4 5
11. pay attention to points I might otherwise have ignored.
    1 2 3 4 5
12. develop the ability to think critically about my assignments.
    1 2 3 4 5
13. develop the ability to think critically about the assignments of my peers.
    1 2 3 4 5
14. assess myself objectively.
    1 2 3 4 5
15. assess my peers objectively.
    1 2 3 4 5
16. follow/understand my progress.
    1 2 3 4 5
17. understand in what I was being assessed.
    1 2 3 4 5
18. organise my work better.
    1 2 3 4 5

I would
19. have liked there to have been more criteria.
    1 2 3 4 5
20. have liked to have done more assignments.
    1 2 3 4 5
21. have liked to have assessed more assignments.
    1 2 3 4 5
22. like to be assessed in this way in all my subjects.
    1 2 3 4 5

PART ‘C’

We request that you write in the space below your comments about the method of assessment, to which you were introduced, trying to give answers to the following questions but giving some ideas which could help in the research:

1. What were the strong points of the method?
2. What were the problems you encountered?
3. What suggestions would you make to improve the method?

____________________________________________________________________________________
____________________________________________________________________________________
____________________________________________________________________________________
____________________________________________________________________________________
____________________________________________________________________________________
____________________________________________________________________________________
____________________________________________________________________________________
Abridged formulation of Antonopoulou’s assessment questionnaire for AARP research purposes:

<table>
<thead>
<tr>
<th>AARP RESEARCH ASSESSMENT QUESTIONNAIRE</th>
<th>1=Completely disagree 2=Tend to disagree 3=Undecided 4=Tend to agree 5=Completely agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I find it an interesting form of assessment.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>2. I understood why this method of assessment was being used.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>3. I found it easy to assess the work of my peers (peer-assessment).</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>4. I found it easy to assess my peers objectively.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>5. I found it easy to assess my own work (self-assessment).</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>6. I found it easy to assess myself objectively.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>The criteria checklists helped me to:</td>
<td></td>
</tr>
<tr>
<td>7. understand the areas in which I was being assessed.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>8. be a more active participant in the lessons.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>9. pay attention to points I might otherwise have ignored.</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>10. develop critical thinking skills.</td>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>
## AARP RESEARCH ASSESSMENT QUESTIONNAIRE – ANSWERS FROM GROUP E

Circle the numbers that represent how you feel about aspects of the assessment.  
1 = Completely disagree  2 = Tend to disagree  3 = Undecided  4 = Tend to agree  5 = Completely agree

1. I find it an interesting form of assessment.

31%, which is the highest percentage of all groups, all years, were undecided about its benefits, while the remaining 69% found it quite interesting.

2. I understood why this method of assessment was being used.

51.70%, which is very much the highest in all years, were undecided, while 41.40% quite understood and only 6.90% very much understood.

3. I found it easy to assess the assignments of my peers (peer-assessment).

20.70%, the highest of all groups, all years, quite disagree that it was easy.

4. I found it easy to assess my peers objectively.

37.90%, which is the second-highest percentage in all years, were undecided, 41.40% quite agreed and 20.70% very much agreed.

5. I found it easy to assess my own assignments (self-assessment).

44.80%, which is the second-highest score in all groups, all years, were undecided about whether it was easy or not, while 31% thought it was quite easy. No one thought it was very easy.

6. I found it easy to assess myself objectively.

41.40%, which is again the highest percentage in any group in any year were undecided, while 37.90% quite agreed it was easy. 17.20% thought it was very easy.

The criteria checklists helped me to:

7. understand the areas in which I was being assessed.

17.20% were undecided, and equal percentages of 41.40% quite agreed and very much agreed.

8. be an active participant in the lessons.

31%, which again was the highest percentage in all groups, all years, were undecided about the usefulness of the lists in this respect, while 44.80% quite agreed and 24.10% very much agreed.

9. pay attention to points I might otherwise have ignored.

6.90% were undecided, 37.90% quite agreed and 51.70% very much agreed, showing their liking for the criteria check-lists.

10. develop critical thinking skills.

20.70% were undecided, 58.60% quite agreed and 17.20% very much agreed.
Section I - Writing – assessment criteria
Please indicate on a scale of 1 to 5, how much you agree with each statement. Please add comments, if and where you feel it is appropriate.

1. For each of the two home writing assignments, you were given the criteria by which you would assess and be assessed in advance of task completion. Do you agree that the five criteria you were given for each one of the tasks contributed to your understanding of the demands of the task?

   Strongly Disagree      Strongly Agree

   1  2  3  4  5

   Please say why …………………………………………………………………

2. Before the course writing homework tasks are set, do you believe that students should help to select the criteria (for the criteria check-list) by which each writing task will be assessed?

   Strongly Disagree      Strongly Agree

   1  2  3  4  5

   Please say why …………………………………………………………………

3. Students were asked to bring the draft copies of their first assignment and have their peers comment on them in class before submitting the final version. Do you believe that it is useful to have other students read and comment on the assignment drafts in class before officially submitting them?

   Strongly Disagree      Strongly Agree

   1  2  3  4  5

   Please say why …………………………………………………………………

4. You had some practice in peer-assessment with five writing task samples, which were compared with the instructor’s assessment. Was this practice with samples before doing the real thing useful?

   Strongly Disagree      Strongly Agree

   1  2  3  4  5

   Please say why …………………………………………………………………

5. Do you think more training of this kind needs to be given for students to be able to assess peers and themselves more accurately?

   Strongly Disagree      Strongly Agree

   1  2  3  4  5

   Please say why …………………………………………………………………
6. Do you believe that students benefit from the process of assessing the completed writing assignments of other students (peer-assessment)?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Please say why ...............................................................................................................

7. Do you feel more confident about assessing the writing tasks of peers because it is done anonymously?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Please say why ...............................................................................................................

8. Do you believe that you and your peers can be trusted to assess your writing assignments accurately and resist the temptation to ‘cheat’ by bumping up grades?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Please say why ...............................................................................................................

9. Do you believe that the training in assessment you had with samples and then with peers enabled you to assess your own writing assignments accurately (self-assessment)?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Please say why ...............................................................................................................

10. Do you believe that self-assessment of writing assignments is a useful exercise?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Please say why ...............................................................................................................

11. Are you happy with the equal role that the instructor played with peers and selves in the writing assessment process?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Please say why ...............................................................................................................

12. You were given back the completed assessments (self-, peer- and instructor-) of both writing assignments. Did this provide sufficient feedback on your writing assignments?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Please say why ...............................................................................................................

47
13. Do you feel that you benefitted from the whole cycle of peer-assessment, self-assessment and instructor-assessment of writing tasks and that it was worthwhile?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

Please say why …………………………………………………………………

Section II – Speaking assessment

14. Do you believe that students’ speaking abilities on Language Mastery I should be assessed by peers as well as the instructor?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

Please say why …………………………………………………………………

15. Do you believe that the five criteria in the criteria check-list, given for oral assessment, contributed to your understanding of the demands of the oral assignment?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

Please say why …………………………………………………………………

16. Do you believe that the criteria check-list, with the five criteria given for the assessment of oral presentations should be created by students themselves in agreement with the instructor?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

Please say why …………………………………………………………………

17. Do you believe that the practice session with the mock-presentations by older students helped you both in your preparation of your oral assignment and in your ability to assess?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

Please say why …………………………………………………………………

18. Do you believe it is more difficult to be objective with peers when assessing their speaking than it is with their writing?

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Strongly Agree</th>
</tr>
</thead>
</table>

Please say why …………………………………………………………………
19. Do you believe that peer-assessment of presentations should be conducted anonymously, with all members of the class assessing, as it was?

| Strongly Disagree | 1 | 2 | 3 | 4 | 5 | Strongly Agree |

Please say why ………………………………………………………………………..

20. Do you believe that students can learn to be objective in assessing their peer and their own oral presentations and avoid the temptation to ‘cheat’ by bumping up grades?

| Strongly Disagree | 1 | 2 | 3 | 4 | 5 | Strongly Agree |

Please say why ………………………………………………………………………..

21. Do you believe that the whole process of peer-assessment and self-assessment of writing and speaking on Language Mastery I has been of value to you in your language learning?

| Strongly Disagree | 1 | 2 | 3 | 4 | 5 | Strongly Agree |

Please say why ………………………………………………………………………..

22. Do you believe that instructors are justified in trusting their students to conduct peer and self-assessment of writing and speaking assignments and in this way contributing to their final course grades?

| Strongly Disagree | 1 | 2 | 3 | 4 | 5 | Strongly Agree |

Please say why ………………………………………………………………………..

Any further comments about the assessment process:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

49
APPENDICES 26a – 26j
Pre-Study Questionnaire Response Analysis
Question 1 – Did you find this form of evaluation interesting?

57.10% in Group A and 60% in Group B found it quite interesting, while 7.10% in Group A and 16% in Group B found it a very interesting type of evaluation. A total of 64.20% in Group A and 76% in Group B were therefore positive about it, as opposed to only a total of 11% in Group A and 12% in Group B who were negative. While 12% in Group B were undecided, more than double that number, with 25% were undecided in Group A.

Student comments:-

A1 – It helps you to learn to grade and judge.
A5 – Through this technique we learn to assess ourselves and others also.
B3 – It was altogether new from anything I had learned previously.
B5 – This technique made the lesson quite interesting.
Question 2 – Did you understand why this method of assessment was being used?
46.40% in Group A and 52% in Group B were quite aware, whereas 14.20% in Group A and 20% in Group B were extremely aware. On the other hand, 28.60% in Group A and 20% in Group B were undecided, while 10.70% in Group A and 4% in Group B, were rather unaware and 0% in either group claimed to be completely unaware.

Student comments:-
A4 – It was evident and clear exactly what we had to evaluate in the written assignments we had in front of us.
A14 – Development of critical thought.
B7 – Objective assessment and a spherical picture of the assignments of fellow students.
B22 – ...we all knew the points on which to base our assessment, so that we understood how we should assess in each case.
Question 3 – How easy was it to assess your peers (peer-assessment)?

In Group A 10.70% found it very easy to peer-assess, while the same percentage of 10.70% found it very difficult. At the same time, 20% in Group B claimed it was very easy, while 0% said it was very difficult. Whereas 28.60% in Group A found peer-assessment quite easy, 48% in Group B found it so. Thus we can see wide discrepancies between Groups A and B at both ends of the scale, with close agreement only amongst those who were undecided, with 32.10% in Group A and 28% in Group B.

Student Comments:-
A5 – It is difficult to correct the scripts of friends because you cannot be objective.
A16 – ...it is difficult to grade a fellow student on the basis of what they really deserve since there are friendships among us.
B9 – ...in the beginning, I did not feel capable of doing such a thing.
B11 – To assess others is not particularly difficult. I have more difficulty in assessing myself.
Question 4 – How easy did you find it to be objective in peer-assessment?

3.60% in Group A and 0% in Group B found it quite difficult, while 7.10% in Group A and 0% in Group B found it very difficult. The rather large percentage of 42.90% in Group A and considerably smaller percentage of 16% in Group B were undecided. On the contrary, only 25.0% in Group A found it quite easy, while a much larger 60% in Group B found it so. Similar percentages of 21.40% in Group A and 24% in Group B found it very easy to be objective in peer-assessment.

Student comments:-
A6 – The criteria give a clear focus to everyone and the evaluation is done more fairly.
A8 – It helped in the objective assessment of my fellow-students, since I would be in the same position and logically it therefore requires some objectivity.
B1 – ...in this way one understands how one is being judged and this must lead to better results in the assignments that follow.
B17 – Naturally, it is more difficult than classic forms of assessment because greater powers of judgement are required in order to assess oneself or one’s fellow-students more objectively.
Question 5 – *How easy was it to assess yourself (self-assessment)?*

7.10% in Group A and 4% in Group B claimed it was very easy to self-assess, while 14.30% and 16% respectively found it quite easy to self-assess. 25% in Group A, as opposed to more than double, with 52%, in Group B were undecided. Group A, with 35.79% found it quite difficult to self-assess and 17.80% found self-assessment very difficult, while only 16% in Group B found it quite difficult and only 12% found it very difficult.

**Student comments:-**

*A11* – It is difficult to assess yourself with these criteria.
*A14* – We do not know if our assessment is correct. We should give a reason for the grade we give.
*B2* – ...in the beginning I considered the existence of a copy of my assignment for assessment by me superfluous. Afterwards, though, I realised its usefulness.
*B16* – The fact that I had to assess myself was quite difficult, since it was not easy to identify the errors I made.
Question 6 – How easy did you find it to be objective in self-assessment?

In Group A, 7.10% did not find it very easy to be objective in self-assessment and the same percentage of 7.10% did not find it easy at all, while 0% did not find it very easy and 4% did not find it easy at all, in Group B. Approximately the same percentages of 42.90% and 44.0% in Groups A and B respectively were undecided, while 35.70% in Group A and 40% in Group B found it quite easy and 7.10% and 12% in Groups A and B respectively found it very easy to be objective in their self-assessment.

**Student comments:**

**A13** – I learned to assess both myself and my classmates more objectively.

**A20** – I had to evaluate myself and my fellow-students objectively, something which I found particularly difficult since I did not want to be unfair to anyone.

**B9** – With time I was more objective and strict.

**B11** – To assess others is not particularly difficult, but I have more difficulty in assessing myself. I try to be objective.
Question 7 – How useful did you find the assessment criteria checklists in helping you to understand the areas in which you were being assessed?

14.30% in Group A and 4.00% in Group B did not find them very helpful, while 7.10% in Group A and 0% in Group B did not find the criteria helpful at all. 25% in Group A and 20% in Group B were undecided. However, 42.90% in Group A and 52% in Group B found them quite helpful, while 10.70% and 24% respectively found them very helpful in this respect.

Student comments:-
A4 – You have the ability to avert mistakes since you know exactly how you will be assessed and so you can focus on that to achieve a better result.
A6 – It helps you understand the way in which the assignment has to be done and the way in which it has to be assessed.
B13 – The student/assessee already knows which points are given particular weight, which helps him/her quite a lot.
B18 – All students were assessed on the basis of the same criteria and on assignments which were based on the same subjects for everyone.
APPENDIX 26h

Question 8 – Were the assessment criteria checklists helpful from the point of view of making you a more active participant in the lessons?

0% in Group A and 4% in Group B found them not very helpful, while 10.70% and 0% respectively did not find them helpful at all. Similar percentages of 25% in Group A and 28% in Group B were undecided, while similar percentages of 46.40% in Group A and 48% in Group B found them quite helpful and 17.80% and 20% in Groups A and B respectively thought they were very helpful.

Student comments:-
A7 – It obliges the students in a very direct way to pay attention to a presentation or to look closely at an assignment.
A17 – I learn through my mistakes and the mistakes of others. I improve. I have more interest in the lesson.
B14 – ...the students have to be attentive and listen to the assignments which are presented and to grade accordingly.
B17 – It helps each person to take an active part in the lesson. I could also refer to it as a form of dialogue or communication which takes place between the teachers and the students.
Question 9 – How useful were the assessment criteria checklists in helping you to note points that you might otherwise have ignored?

7.10% in Group A did not find the checklists very helpful in this respect, while 3.60% did not find them helpful at all, while 4% in Group B did not find them very helpful and 0% found them not at all helpful. 17.80% in Group A were undecided about their usefulness, while 4% in Group B were the same. 42.90% in Group A and 52% in Group B found them quite helpful, while 28.60% in Group A and 40% in Group B found the criteria very helpful indeed in this respect.

Student comments:-
A6 – I notice mistakes which I didn’t notice in the beginning of my assignments.
A10 – I realised that it was a good and helpful system, since it gives you a focused view of the subject and the criteria with which you are assessed.
B15 – The strong points of this technique were having to listen carefully to others and to identify mistakes which it would be better to avoid yourself.
B20 – I learned to pay attention to points that I would otherwise have missed.
Question 10 – *Do you believe that the assessment criteria checklists helped in the development of critical thinking skills?*

3.60% in Group A and 4% in Group B did not find the criteria very helpful in this respect, while 7.10% and 0% in Groups A and B respectively, did not find them helpful at all. 35.70% in Group A and 28% in Group B were undecided concerning their helpfulness, while 50% in Group A and 48% in Group B found them quite useful, and only 3.60% in Group A as opposed to 20% in Group B found them very useful in developing critical thinking skills.

**Student comments:**

*A12* – This method helped in the development of critical thinking.

*A18* – The strong points in this method are the way it gives us the opportunity to develop critical abilities and thinking. The brain remains alert and creates an objective view of things.

*B2* – The strong points were that I could get to know myself better and as an extension of that my strong points and my weaknesses. In that way it helps me to improve with each assignment and each lesson.

*B17* – ...it is more difficult than classic forms of assessment because greater powers of judgement are required in order to assess objectively.
APPENDICES 27a – 27j
Main Study Questionnaire Response Analysis
Question 1 – Did you find this form of evaluation interesting?

Only Groups G and H in 2008-2009, with scores of 3.80% and 5% respectively, thought it was only slightly interesting. 4.80% in Group F 2007-2008 and none of the remaining groups thought it was not at all interesting. Those undecided ranged from 0%, in the case of Group D 2006-2007, to similar scores of 30.80 from Group G, 2008-2009 and 31.00% from Group E, 2007-2008. Those who found it quite interesting ranged from 53.80% for Group G, 2008-2009 to 76.90% in Group C, 2006-2007. In Group E, 2007-2008, 0% found this form of assessment very interesting, with percentages ranging from 9.50% for Group F, 2007-2008 to a maximum of 31.25% for Group D, 2006-2007.

Student comments:-

D16 – It is quite an interesting method, which helps students in a positive way.
E14 – The lesson becomes interesting and the method helps a lot with our ability to be able to perform in the future as teachers.
F14 – It was very interesting and I found it particularly helpful from the aspect of becoming more objective, both in my self-assessment and in peer-assessment.
G25 – With this original method of assessment, the strong points and the weak points of the students could be assessed better than with the traditional end-of-semester examination.
APPENDIX 27b

Question 2 – Did you understand why this method of assessment was being used?

No groups were unable to understand why this method of assessment was being used, while 0% in both groups during the years 2006-2007 and 2007-2008 and 3.80% and 15% in Groups G and H in 2008-2009 could understand only slightly. A range of 10% in Group H, 2008-2009 to 51.70% in Group E, 2007-2008, were undecided. Groups D and E, 2007-2008 had similar values of 41.40% and and 42.80%, while Groups F and G in 2008-2009, along with Group D, 2006-2007 had similar values of 50%, with a slightly higher value of 53.80% in Group C, 2006-2007, who understood quite well why this type of assessment was being used. Percentages for those who understood its use very well ranged from 6.90% and 14.30% in Groups E and F, 2007-2008 respectively, with similar values of 26.90% and 25% in Groups G and H, 2008-2009 and 23% in Group C, 2006-2007. Group D, 2006-2007, best understood its usefulness with a value of 37.50%.

Student comments:-

E12 – The assessment conducted this way is more accurate and it helped me to understand on what exactly I would be examined.

F16 – Assessment was immediate and quite successful, even if it was difficult to assess ourselves.

G23 – The strong point of the method was the objective assessment and the joint contribution of opinions between peers, instructor and myself.

H3 – You know exactly what the other is looking for when they grade your script. In this way you pay attention to details which you otherwise might not have done. So you improve.
Question 3 – *How easy was it to assess your peers (peer-assessment)?*

Noone in the three groups, Group D, 2006-2007, Group F, 2007-2008 and Group G, 2008-2009 found it not at all difficult, while the three remaining groups had low values between 3.40% and 7.70%. Between 3.80% in Group G, 2008-2009 and a maximum of 20.70% in Group E, 2007-2008 found peer-assessment quite difficult, while values between 28.60% in the latter group and 53.80% in Group C, 2006-2007 showed that all groups were not sure. 23% in Group C, 2006-2007, found it quite easy, while between 40% to 47.60% in the other five groups found it so. Only 7.70% in Group G, 2008-2009 found it very easy, as did 9.50% in Group F, 2007-2008, while noone in all the remaining groups found it very easy.

**Student comments:**

*C3* – I consider it to be quite a good method, but perhaps we should have other criteria so that our correcting could be more accurate.

*D1* – I had difficulty, of course, in assessing others, but it was a very good thing to do.

*E20* – The difficult thing was to recognise the mistakes (non-grammatical) of others, more to do with how they expressed themselves and suggest other alternatives.

*H12* – I realised how difficult it is to evaluate written assignments.
APPENDIX 27d

Question 4 – How easy did you find it to be objective in peer-assessment?

Noone from any of the six groups found it very difficult to be objective. Noone in either Group D, 2006-2007 or in Group E, 2007-2008 found it quite difficult to be objective, while low percentages of between 3.80% (Group G, 2008-2009) and 7.70% (Group C, 2006-2007) in the other four groups found it so. Between 18.75% (Group D, 2006-2007) and 38.50% (Group C, 2006-2007) in all groups were undecided on this issue, while between 41.40% (Group E, 2007-2008) and 60% (Group H, 2008-2009) in all groups found it quite easy and between 7.70% (Group C, 2006-2007) and 31.25% (Group D, 2006-2007) in all groups found it very easy to be objective in their assessment of peers.

Student comments:-
C9 – In the beginning it was difficult to assess ourselves and our classmates, and if we didn’t give a good grade, we felt bad.
D8 – I had never found myself before in similar circumstances and I confess that it caused me anxiety that I would perhaps assess someone more strictly than I should.
E14 – We learn how to assess with impartiality.
H6 – ...the assessment of the assignments of my classmates helped me to begin to judge them objectively too, as I will have to do when I become an English teacher.
Question 5 – *How easy was it to assess yourself (self-assessment)?*

Group C, 2006-2007, with the highest of 39% did not find self-assessment easy at all, while Groups D, 2006-2007 and F, 2007-2008 felt more confident, with 0%. Groups E, in 2007-2008 and G, in 2008-2009 have similar values of 3.40% and 3.80%, while Group H, 2008-2009, in contrast, has the much higher value of 20%. 15.40% in Groups C in 2006-2007 and E in 2008-2009 found self-assessment quite difficult, while the remaining groups ranged between 15.40% and 33.30%. Quite a large percentage in all groups, of 30.80% to 50.00%, were not sure if they found it easy or not, while between 7.70% in Group C, 2006-2007 and a maximum of 38.50% in Group G, 2008-2009, found it quite easy. 0% in Group C, 2006-2007, Group E, 2007-2008 and Group H, 2008-2009 claimed to find it very easy, while between only 6.25% and 11.50% did so in the three remaining groups.

*Student comments:*

**C4** – I learned to be more objective with myself, although it was hard to grade myself.

**D3** – The problem was that I couldn’t really assess myself. I found everything perfect!

**E12** – The assessment conducted this way is more accurate and it helped me to understand on what exactly I would be assessed.

**H5** – It was equally difficult to self-assess since I had to assess and judge using objective criteria.
Question 6 – How easy did you find it to be objective in self-assessment?

10% in Group H, 2008-2009 thought that it was very difficult to be objective while no one in the five remaining groups thought so. Those groups that thought objectivity was quite difficult ranged between 0.00% in Group D, 2006-2007 and 30.80% in Group C, 2006-2007. Only 9.50% in Group E, 2007-2008 were unsure about their objectivity in self-assessment while in the other groups this ranged between 30.80% (Groups C, 2006-2007 and G, 2008-2009) and 41.40% (Group E, 2007-2008). Between 23% (group C, 2006-2007) and 42.80% (Group F, 2007-2008) in all groups found it quite easy to be objective in self-assessment, while between 10% (Group H, 2008-2009) and 31.25% (Group D, 2006-2007) in all groups found it very easy to be objective.

Student comments:-
D4 – I had never assessed with such criteria before and thus the whole concept was a little complicated.
E26 – I do not believe that a person can be sufficiently objective with himself.
F17 – We all assess ourselves very leniently, and our peers quite strictly.
G11 – I could not assess myself objectively.
Question 7 – How useful did you find the assessment criteria checklists in helping you to understand the areas in which you were being assessed?

Noone in the six groups of the Main Study found the checklists not at all useful in this respect. 15% in Group H, 2008-2009 did not find the checklists particularly useful, while in the remaining five groups there was noone who thought so. Between 6.25% (Group D, 2006-2007) and 33.30% (Group F, 2007-2008) in all groups were undecided about their usefulness in this respect, while between 37.50% (Group D, 2006-2007) and 73% (Group G, 2008-2009) in all groups found them quite helpful and between 14.30% (Group E, 2007-2008) and 56.25% (Group D, 2006-2007) in all groups found the checklists very helpful in this respect.

Student comments:-

C4 – I learned to be more objective with myself, although it was hard to grade myself.

D2 – The criteria descriptions in the assessments were exact and very organised. You referred to the exact areas on which you had to assess and be assessed.

E11 – There were very clear criteria for assessment, so that we could judge better ourselves and our fellow-students on these specific areas.

H7 – This assessment particularly helps students to understand the points on which they need to turn their attention during their correcting and assessment of a written assignment.
APPENDIX 27h

QUESTION 8

Question 8 – Were the assessment criteria checklists helpful from the point of view of making you a more active participant in the lessons?

Between 15% and 33.30% thought the criteria checklists were very helpful, between 28.60% and 77%, with the latter high percentage recorded by Group C, 2006-2007, thought they were quite helpful. Noone in Group C, 2006-2007 was undecided, while those unsure in the other 5 groups ranged between 12.50% in Group D, 2006-2007 and a maximum of 31% in Group E, 2007-2008. In Groups C, 2006-2007, E, 2007-2008 and G, 2008-2009, noone thought the checklists were not really helpful, while between 5% and 9.50% felt the same in the remaining three groups. 5% in Group H, 2008-2009 did not find them helpful at all, while noone in the remaining five groups found them so.

Student comments:-

D5 – Improvement concerning the stimulus for students and myself to be active participants.

F5 – It helps the student to take an active part in the lesson. We understand which points we have to pay attention to in the written assignments and in the oral presentation.

G22 – It led most students to participate more actively.

H18 – The strong points of the method were that I knew on what criteria the assignments were being graded. In addition, I took an active part in the lesson and I was not bored.
Question 9 – How useful were the assessment criteria checklists in helping you to note points that you might otherwise have ignored?

All groups found the checklists very useful, with values ranging between 23.80% in Group F, 2007-2008 and 68.75% in Group D, 2006-2007, and quite useful, with values ranging between 25.00% in Group D, 2006-2007 and Group F, 2007-2008, with 61.90%. No one in Group C, 2006-2007 was unsure about their usefulness, while there were between 6.25% (Group D, 2006-2007) and 26.90% (Group G, 2008-2009) unsure in the remaining five groups. Small percentages of 5% in Group H and 7.70% in Group G, 2008-2009, found the checklists of relatively little use, while none of the remaining groups found them so. 3.40% in Group E, 2007-2008 and 5% in Group H, 2008-2009 did not find the checklists at all useful in this respect, while none in the remaining groups found them so.

Student comments:

C8 – This method helped me to pay attention to particular points, which I might not have noticed at all.
D4 – I became familiar with new assessment criteria which I had not thought of until now.
E24 – It helped me to focus on particular points and to understand what was of greater importance.
H15 – It enabled me to notice points otherwise difficult for me to notice.
Question 10 – Do you believe that the assessment criteria checklists helped in the development of critical thinking skills?

All of the groups thought the checklists were very useful in developing critical skills on a range between 7.70% (Group C, 2006-2007) and 37.50% (Group D, 2006-2007) while all groups also thought they were quite useful in this respect, on a wider range, between 34.60% (Group G, 2008-2009) and 77% (Group C, 2006-2007). Between 6.25% (Group D, 2006-2007) and the much higher 38.50% (Group G, 2008-2009) in all groups were undecided about the usefulness of the criteria checklists in this respect. 3.80% in Group G, 2008-2009 and 0% in all other groups thought that the criteria checklists were not altogether useful, while 5% in Group H, 2008-2009 and 3.40% in Group E, 2007-2008 and none in the remaining four groups thought the criteria checklists were not at all useful in this respect.

Student comments:

C1 – This method played an important role in the development of critical thinking regarding assignments.

D8 – We cultivate our critical thinking and experience, something which will be of use to us in our future years as teachers of the English language.

E29 – The good point is that you develop and cultivate critical thinking through assessing others and yourself.

F11 – I developed critical thinking related to the assignments of my fellow-students and I have gradually learned how to assess myself.
APPENDICES 28a-28v
Post-Study Questionnaire Response Analysis
Question 1 – Did the assessment criteria checklists contribute to your understanding of the demands of the task?

3.40% in Group I and 0% in Group J did not find the checklists altogether helpful, while there was no one in either group who did not find them helpful at all. 20.70% in Group I and 9% in Group J were undecided about their usefulness, while 41.40% and 54.50% in Groups I and J respectively found them quite helpful and, at the same time, similar numbers of 34.50% and 36.40% in the same groups, found them very helpful.

Student comments:
I 4 – Yes, because they were very specific.
I 9 – They were very helpful.
J 20 – The criteria matched the task given.
J 26 – ...for the first task it was rather confusing, for the second one it helped.
Question 2 – Do you think that students should contribute to selecting the criteria used for assessment purposes in each writing assignment?

3.40% in Group I and 9% in Group J did not agree at all concerning contributing to the criteria checklists, while 0% and 13.60% in the same groups did not altogether agree. Around the same percentage of 37.90% and 36.40% in Groups I and J respectively were undecided, while 31% in Group I and 27.30% in Group J agreed quite strongly, whereas double the percentage in Group I, with 27.60% compared to 13.60% in Group J strongly agreed.

Student comments:
I 3 – I think that in the 1st year the children aren’t still ready for this action.
I 16 – Yeah, it should be fair.
J 14 – I don’t think the students have the appropriate knowledge to select the criteria.
J 20 – It’s up to the tutor to decide on such matters.
Question 3 – Do you believe it is useful to have other students read and comment on the assignment drafts in class before officially submitting them?

6.90% in Group I and 9.0% in Group J did not agree at all, while 3.40% and 4.50% in the same groups, did not quite agree. Only 3.40% in Group I were undecided, while a considerably greater 22.70% in Group J could not decide one way or the other. 44.80% and 54.50% in Groups I and J respectively were quite in agreement, while 41.40% in Group I and a much smaller 9% in Group J were very much in agreement.

Student comments:-
I 3 – Yes, because let us see how our classmates think.
I 13 – Different opinions might confuse the student.
J 3 – It would be useful if everybody knew the language properly and were unbiased.
J 14 – A fellow student’s opinion helps improvement.
Question 4 – Did the mock assessment of five writing samples, which were compared with the instructor’s assessment, provide useful practice?

There was no one in either group who did not agree with this at all, though there were 3.40% and 4.50% in Groups I and J who were quite in disagreement. 10.30% in Group I and 18.20% in Group J were undecided on this matter. Similar percentages of 51.70% and 50.00% in Groups I and J were quite in agreement, while 34.50% and 27.30% likewise were very much in agreement as to its usefulness.

**Student comments:**

I 5 – Yes, in order to see how different is the point of view of the instructor and the peers.

I 24 – It was helpful in understanding the use of criteria.

J 18 – Yes, because we gradually learn how to assess.

J 24 – I didn’t find it very useful because I believe that everyone has its own starting-point in evaluating others’ work and it is different for every student.
Question 5 - Do you think more training of this kind needs to be provided to enable students to assess peers and themselves more accurately?

6.90% in Group I and 13.60% in Group J did not quite agree, while 0% and 4.50% in the same groups did not agree at all. 24.10% in Group I, and a much higher 40.90% in Group J, were undecided, while 37.90% in Group I and 31.80% in Group J were quite in agreement, compared with 31% in Group I and 9% in Group J who were very much in agreement.

Student comments:
I 7 – Yes, in order to gain experience in doing this and doesn’t seem to them something unfamiliar.
I 11 – Yes, because we are not already well trained.
J 3 – I do not think that needs more training.
J 25 – That was really enough.
Question 6 – Do you believe that students benefit from the process of assessing the completed writing assignments of other students (peer-assessment)?

6.90% in Group I and 9% in Group J quite disagreed with its being beneficial, while none in either group completely disagreed about its benefits. Around the same percentage of 24.10% and 22.70% in Groups I and J respectively were undecided, while 41.40% and 54.50% in the same groups were quite in agreement. Students who were very much in agreement concerning the benefits were more than double in Group I, with 27.60%, compared with 13.60% in Group J.

Student comments:

I 8 – No, they could benefit more by doing other exercises in class, like conversasing in English.
I 19 – [It offers] self-knowledge, expansion of our vocabulary.
J 1 – They gain the experience of assessment.
J 16 – It helps them get a general idea.
Question 7 – Do you feel more confident about assessing the writing tasks of peers because it is done anonymously?

0.0% in Group I and 4.50% in Group J completely disagree, while 3.40% and 4.50% tend to disagree. 3.40% in Group I and 9.0% in Group J are undecided, while 48.30% and 36.40% in Groups I and J tend to agree and similar percentages of 44.80% and 45.40% respectively, very much agree.

Student comments:

I 5 – Anonymously or not, it’s not an issue for me.
I 11 – Yes, because I wouldn’t want to give a bad mark to some of my friends.
J 2 – You keep your identity secret so as to avoid arguments.
J 15 – It has to be done anonymously.
**Question 8 – Do you believe that you and your peers can be trusted to assess writing assignments accurately and resist the temptation to ‘cheat’ by bumping up grades?**

17.20% in Group I and 9% in Group J tended to disagree, while 13.80% and 9% respectively, in the same groups, strongly disagreed. 27.60% in Group I and 36.40% in Group J were undecided, while similar percentages of 34.50% and 36.40% in Groups I and J were quite strongly in agreement, compared with only 6.90% and 9% in the same groups who were strongly in agreement.

**Student comments:**

*I 7 – No, I think it’s difficult.*

*I 11 – Yes, because we want things to be objective. Though, sometimes the grades are not good.*

*J 10 – Cheating may occur.)*

*J 19 – Trust is everything.*
Question 9 – Do you believe that the training in assessment you had with samples and then with peers enabled you to assess your own writing assignments accurately (self-assessment)?

6.90% in Group I and 4.50% in Group J completely disagree, while 13.80% and 22.70% respectively tend to disagree. 31% in Group I and 40.90% in Group J are undecided on this issue, while 34.50% and a much smaller 13.60% tend to agree, while 13.80% in Group I and 18.20% in Group J completely agree.

Student comments:
I 1 – [We gained] experience in evaluating. The criteria have been clarified.
I 3 – Yes, it made me think better.
J 4 – Of course, you gain experience.
J 24 – It’s always hard to assess yourself.
Question 10 – Do you believe that self-assessment of writing assignments is a useful exercise?

17.20% in Group I and 13.60% in Group J did not find it very useful, while 6.90% in Group I and 0% in Group J did not find it useful at all. 20.70% in Group I were undecided, while over double this percentage, with 45.40% in Group J were uncertain of its usefulness. On the other hand, 41.40% in Group I and 31.80% in Group J thought it was quite useful, and 13.80% and 9% in the same groups thought it was very useful.

Student comments:
I 9 – No, we would benefit more through the dialogue.
I 12 – No, because I would hardly ever mark myself badly, to tell the truth.
J 3 – We practise on what we need to learn so as to become better.
J 5 – Yes and no.
Question 11 – Are you happy with the equal role that the instructor played with peers and selves in the writing assessment process?

6.90% in Group I and 4.50% in Group J were not quite in agreement, while 6.90% and 0% in the same groups were not in agreement at all. 17.20% in Group I and 22.70% in Group J were uncertain, while 44.80% and 36.40% in the same groups were quite in agreement, and 24.10% and 36.40% respectively were very much in agreement.

Student comments:-

I 8 – Yes, because it gave us the opportunity to express our opinion about ourself and our classmates too.

I 12 – Yes, because this is how things should be generally.

J 16 – Quite happy.

J 26 – I think that the role of the instructor should be slightly stronger, because he is more educated and experienced.
Question 12 – Did the completed assessment (self-, peer- and instructor-) of both writing assignments provide you with sufficient feedback?

No one completely disagreed on this matter, while 3.40% in Group I and 4.50% in Group J rather disagreed. 20.70% in Group I and 36.40% in Group J were undecided, while 41.40% and 31.80% respectively were quite in agreement, whereas 34.50% in Group I and 27.30% in Group J were completely in agreement.

Student comments:-
I 5 – It proved how most peers have a common way of thought.
I 19 – It was very helpful.
J 6 – I think we can improve ourselves through this process.
J 19 – We need comments and suggestions more than grades.
Question 13 – Do you feel that you benefitted from the whole cycle of peer-assessment, self-assessment and instructor-assessment of writing tasks and that it was worthwhile?

No one in either group completely disagreed about the benefits, though 6.90% in Group I and 4.50% in Group J tended to disagree. 20.70% in Group I and 27.30% in Group J were undecided, while similar percentages of 48.30% and 50% respectively tended to agree, while 24.10% in Group I and 18.20% in Group J completely agreed.

Student comments:-

I 1 – Being able to assess plays a more important role than people think in the learning procedure.
I 11 – Of course, because I made a good practice of my English.
J 2 – Of course, yes, because it is part of the course and we learn more by practising.
J 26 – I think it was good experience of new methods. However, I think it took more time than necessary.
Question 14 – Do you believe that students’ speaking abilities on LM I should be assessed by peers as well as the instructor?

No one completely disagreed on this issue, though 31% in Group I and 13.60% in Group J tended to disagree. 24.10% and 18.20% in Groups I and J respectively were undecided, whereas 24.10% in Group I and a much higher 50% in Group J, tended to agree while 20.70% and 18.20% respectively very much agreed.

Student comments:
- I 5 – Assessed yes – but no equal role included.
- I 13 – It is not necessary.
- J 6 – It would be a good idea.
- J 19 – No because peers are not perfect speakers.
Question 15 – Do you believe that the five criteria in the criteria checklist contributed to your understanding of the demands of the oral assignment?

3.40% in Group I and 9% in Group J rather disagreed, while noone in either group completely disagreed. 13.80% in Group I and 18.20% in Group J were undecided, while 44.80% and 40.90% respectively, were quite in agreement. 37.90% in Group I and 31.80% in Group J were very much in agreement about the usefulness of the checklists, in this respect.

Student comments:-
I 4 – Yes, they were addressing to our skills.
I 11 – Very much, because I knew what I was grading.
J 10 – Strongly agree for that reason.
J 16 – The criteria was sufficient.
Question 16 – Do you believe that the criteria check-list, with the five criteria for the assessment of oral presentations should be created by students themselves in agreement with the instructor?

13.80% in Group I and 9% in Group J were not quite in agreement, while 13.80% and 18.20% in the same groups were not at all in agreement. 20.70% in Group I and almost double that percentage, with 40.90% in Group J, were uncertain. On the other hand, 34.50% in Group I and 22.70% in Group J were quite in agreement concerning students contributing to criteria, while 17.20% and 9% in the same groups were very much in agreement.

Student comments:-

I 8 – This would be a nice idea! Because it cultivates collaborations between us and the instructor.
I 13 – No, you have the experience. We’re not so mature yet.
J 16 – I wouldn’t find this useful.
J 26 – ...it should be the instructor to have the leading role in setting out the criteria.
Question 17 – Do you believe that the practice session with the mock-presentations by older students helped you both in your preparation of the oral assignment and in your ability to assess?

No one in either Group I or J completely disagreed, though 3.40% and 4.50% tended to disagree. 10.30% in Group I and a much higher 36.40% in Group J were undecided. 24.10% in Group I tended to agree as did 36.40% in Group J. 62.10% in Group I and the very much lower 22.70% in Group J were completely in agreement.

Student comments:-
I 8 – I learned some things.
I 11 – No, because we weren’t all very good and generally they made us anxious.
J 2 – Yes, coz we could take ideas from older students who are more experienced than we.
J 16 – It could have been done better.
Question 18 – Do you believe it is more difficult to be objective with peers when assessing their speaking than it is with their writing?

10.30% in Group I and 4.50% in Group J are very much in disagreement, while 6.90% and 9.0% respectively tend to disagree. 10.30% in Group I and 36.40% in Group J are undecided, while 24.10% in Group I and 36.40% in Group J tend to agree and the very large 62.10% in Group I and 22.70% in Group J very much agree.

Student comments:
11 – Almost the same.
111 – No, it is easier I think!
16 – It is certainly harder to be objective.
26 – I don’t find big difference because they are two different kinds of tasks so they require different ways of assessing, but the level of objectivity will remain the same.
Question 19 – Do you believe that peer-assessment of presentations should be conducted anonymously, with all class members assessing, as it was?

No one in either group completely disagreed, while 3.40% in Group I and 13.60% in Group J tended to disagree. 6.90% in Group I and 9% in Group J were undecided, while 34.50% and 45.40% respectively tended to agree, whereas 55.20% in Group I and 31.80% in Group J completely agreed.

Student comments:

I 3 – Yes, in order to be expressed more freely.
I 20 – Yes, anonymously it’s better.
J 3 – I do not think that anyone cares for being anonymously or not.
J 16 – I strongly agree.
Question 20 – Do you believe that students can learn to be objective in assessing their peer and their own oral presentations and avoid the temptation to ‘cheat’ by bumping up grades?

6.90% in Group I and 0% in Group J completely disagreed, while 6.90% and 9.0% respectively tended to disagree. 27.60% and 22.70% respectively were undecided, while 41.40% in Group I and 35.40% in Group J tended to agree, whereas 17.20% in Group I and 31.80% in Group J completely agreed.

Student comments:
I 2 – Yes, it’s an effective way to learn not to be subjective.
I 8 – It is hard enough to avoid temptation in our first year of studying.
J 1 – They are always more subjective with their own assessment.
J 23 – Sometimes not even teachers are subjective.
Question 21 – Do you believe that the whole process of peer-assessment and self-assessment of writing and speaking on Language Mastery I has been of value to you in your language learning?

20.70% in Group I and 31.80% in Group J were undecided, while 6.90% and 4.50% respectively were in slight disagreement and 6.90% in Group I and 0% in Group J completely disagreed about its usefulness. On the other hand, similarly high percentages of 41.40% in Group I and 45.40% in Group J were quite in agreement about its value, while 24.10% and 18.20% respectively, were completely in agreement about its usefulness in their language learning.

**Student comments:**

I 3 – No, more homeworks in class and at home would be more effective.
I 11 – Of course! It helped me improve my English.
J 3 – They help us to learn more.
J 16 – It has been invaluable.
Question 22 – Do you believe that instructors are justified in trusting their students to conduct peer- and self-assessment of writing and speaking assignments and in this way contributing to their final grades?

10.30% and 4.50% in Groups I and J respectively completely disagree, while 6.90% and 0% in both cases tend to disagree. 13.80% in Group I and a much larger 36.40% in Group J are undecided on this issue. 58.60% in Group I and 50% in Group J tend to agree on this issue, while similar percentages of 10.30% and 9% respectively completely agree.

Student comments:-

I 10 – Not all students are objective.
I 20 – Students can’t judge objectively while assessing themselves or their peers.
J 3 – Yes, of course.
J 16 – Instructors should trust their students.
BIBLIOGRAPHY


MUSICOGRAPHY