

Bond University

DOCTORAL THESIS

Predicting academic performance at university and investigating the validity of a post-entry language assessment (PELA)

Lydster, Cameron

Award date:
2024

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.



**Predicting Academic Performance at University and
Investigating the Validity of a Post-entry Language
Assessment (PELA)**

Cameron Lydster

Submitted in total fulfilment of the requirements of the degree
of Doctor of Philosophy

February 2024

Faculty of Society and Design

Dr Beata Webb, Dr Masanori Matsumoto, and Dr Gaelle Brotto

*This research was supported by an Australian Government Training Program
scholarship.*

ABSTRACT

The landscape of higher education has changed considerably due to internationalisation and massification. Although a positive phenomenon, challenges concerning academic language proficiency have arisen. Concomitantly, rates of attrition have become concerning for all stakeholders; therefore, predictive studies identifying students likely to require additional support have been undertaken to enhance retention and success rates. Similarly, post-entry language assessments or PELAs assist in identifying students who may lack academic language skills. PELAs are ubiquitous in Australia, yet there is limited literature investigating the validity of PELA score interpretations and uses and the potential predictive qualities of the measures. This thesis aimed to determine whether academic and non-academic factors could predict academic performance (i.e., grade point average, [GPA]) of undergraduate students. The main factor investigated was Bond English Language Assessment (BELA), defined as a PELA screening tool measuring academic essay writing skills of undergraduate students. Via a mixed-method approach, this thesis also aimed to investigate the validity of the interpretation and use of BELA scores.

Findings provided support for the validity argument for BELA in the context of its use at the research site. The research identified concerns regarding inter-rater reliability, academic integrity, and the provision of feedback. Recommendations in response to each concern, included more in-depth and frequent rater training for academic staff, with the development of an online rater training module, and routinely monitoring ratings. It was also suggested that further investigation be undertaken concerning the administration of BELA in an environment which allows invigilation, especially with the rapid development of generative artificial intelligence. Furthermore, it was recommended that all students receive general, online feedback on their essays. A final recommendation was to determine the appropriateness of widening the group of students required to attend Academic Skills Centre (ASC) consultations.

Turning to the main purpose of the research, via logistic regression, this thesis identified academic and non-academic factors predicting academic performance of undergraduate students. Regarding students likely to fail after two semesters, Indigenous Australian students were 4.06 times more likely to fail in comparison to non-Indigenous students; students with Below satisfactory academic writing were 3.77 times more likely to fail compared to students with Satisfactory writing; Bond University College students were 3.12 times more likely to

fail than students from other faculties; students who did not attend ASC were 2.89 times more likely to fail in comparison to those who engaged; males were 1.93 times more likely to fail compared to females, English as an additional language or dialect (EAL/D) students were 1.89 more likely to fail in comparison to English-speaking background students; finally, Bond Business School students were 1.74 more likely, compared to students from other faculties, to fail after two semesters.

In terms of high academic achievement, the odds of maintaining a GPA above 3.0 (i.e., a Distinction average or above) after two semesters were: 5.16 times higher for students with Satisfactory academic writing, 4.63 times greater for Health Sciences and Medicine students, 1.67 times higher for English-speaking background students, 1.45 times greater for females, and 1.03 times more for older students.

The creation of a profile of students at risk of potential academic failure and, conversely, a profile of students likely to excel academically, assists decision makers in allocating resources to support students in an effort to reduce attrition and enhance retention and success rates. This thesis contributes to the literature on success and retention at university and evaluating the validity of PELAs. This research opens the door to further studies into the role of predictive factors, including PELAs, within other contexts.

KEYWORDS

Undergraduate university students, academic performance, attrition, retention, success, post-entry language assessment, PELA, validity.

DECLARATION BY AUTHOR

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy (PhD) by Research.

I declare that the research presented within this thesis is a product of my own original ideas and work, and contains no material which has previously been submitted for a degree at this university or any other institution, except where due acknowledgement has been made.

Name: Cameron Lydster

Signature:

Date: February 14th, 2024

PUBLICATIONS AND CONFERENCE PRESENTATIONS

JOURNAL PUBLICATIONS

- Lydster, C., & Murray, J. (2019). "Not just a tutor": Successful supplementary tuition for Australian Indigenous students in higher education. *Journal of Academic Language and Learning*, 13. A140-A160. Retrieved from <https://journal.aall.org.au/index.php/jall/article/view/609/435435455>
- Lydster, C., & Murray, J. (2018). Understanding the challenges, yet focusing on the successes: An investigation into Indigenous university students' academic success. *The Australian Journal of Indigenous Education*, 48(2). 107-118. <https://doi.org/10.1017/jie.2018.15>
- Lydster, C., & Brown, S. (2017). The value of Post-Entry Language Assessment (PELA): Outcomes from a first semester undergraduate subject. *Journal of Academic Language and Learning*, 11. A1-A19. Retrieved from <https://journal.aall.org.au/index.php/jall/article/view/434>

CONFERENCE PRESENTATIONS

- Lydster, C. (2022). Direct Entry Programs: Identity crisis or turning point? *English Australia Pre-Conference Workshops*. 8th September 2022.
- Lydster, C. (2022). Post-entry language assessment (PELA): An overview. *English Australia Post-entry Academic Language special interest group*. 7th April, 2022.
- Lydster, C. (2021). Archetypal support mechanisms in post-entry English and academic language (drop in and one-on-one consultations). *English Australia Pre-Conference Workshops*. 13-17th September 2021.
- Lydster, C. (2019). Bond English Language Assessment (BELA). The Applied Linguistics Conference 2019: *Applied Linguistics and Language Teaching: Making Connections*. Edith Cowan University, 25-27th November 2019.
- Lydster, C., & Brown, S. (2015). Investigating the efficacy of a PELA at Bond University. *Critical Intersections: the 12th Biennial Conference of the Association for Academic Language and Learning*. University of Wollongong, 25 -27th November 2015.

ETHICS DECLARATION

The research associated with this thesis received ethics approval from the Bond University Human Research Ethics Committee. Ethics application number CL03283.

COPYRIGHT DECLARATION

No copyright declaration.

No published manuscripts were included for publication within this thesis.

ACKNOWLEDGEMENTS

This project has taken close to seven years to complete, and I am indebted to so many individuals who have encouraged, supported, motivated and kept my feet on the ground. First of all, I would like to express my sincere gratitude to my supervisory team, Dr Beata Webb, Dr Masanori Matsumoto, and Dr Gaelle Brotto. I have learnt so much from you these past seven years, and I am so glad I was able to make it this far. Your feedback always motivated me to continue, and I appreciate your kind words throughout the process, especially when the project was not completely smooth sailing. I would also like to thank Dr Pamela Humphreys, who provided much guidance early on and convinced me to learn more about language assessment validity; I am so glad I did. To Dr Alan Finch, you planted the seed and made me realise this endeavour was a possibility. Thank you for showing faith in me and being an inspiration. To Dr Raoul Mortley, I appreciate all of your support and encouragement, particularly at the early stages of the project.

To my wonderful colleagues at Bond University who have always encouraged me to keep going, thank you! In particular, thank you to my workmates at Academic Skills Centre. You have been an ideal sounding board throughout the process, and much of the research conducted is thanks to your hard work. I would also like to thank the many students who have motivated me to continue this project over the years.

Most importantly, I would like to thank my family for supporting me throughout this research journey. To my parents, Alan and Anne, for providing me with outstanding educational opportunities and encouraging me to achieve my goals. To Mika, Kiara, and Jack, thank you for being patient with me and being my motivation to complete this. This is for you. *Fortius Quo Fidelius.*

TABLE OF CONTENTS

TITLE PAGE.....	i
ABSTRACT.....	ii
KEYWORDS.....	iv
DECLARATION BY AUTHOR.....	v
PUBLICATIONS AND CONFERENCE PRESENTATIONS.....	vi
JOURNAL PUBLICATIONS	vi
CONFERENCE PRESENTATIONS	vi
ETHICS DECLARATION	vii
COPYRIGHT DECLARATION	viii
ACKNOWLEDGEMENTS.....	ix
TABLE OF CONTENTS.....	x
LIST OF TABLES.....	xvi
LIST OF FIGURES	xviii
ABBREVIATIONS	xxii
CHAPTER 1: INTRODUCTION.....	1
1.0 OVERVIEW OF THE STUDY AREA	1
1.1 THE EVOLVING CONTEXT AND CHALLENGES WITHIN HIGHER EDUCATION.....	1
1.2 POST-ENTRY LANGUAGE ASSESSMENT	3
1.3 RESEARCH OVERVIEW	4
1.4 SIGNIFICANCE OF THE CURRENT INVESTIGATION	4
1.5 OUTLINE OF CHAPTERS.....	5
CHAPTER 2: PREDICTING ACADEMIC PERFORMANCE.....	7
2.0 INTRODUCTION	7
2.1 BACKGROUND ON ACADEMIC PERFORMANCE.....	7

2.2 THE PROBLEM OF ATTRITION, RETENTION AND SUCCESS.....	8
2.3 PREDICTING PERFORMANCE IN RESPONSE TO THE PROBLEM.....	11
2.4 OVERVIEW OF COMMONLY CITED PERFORMANCE PREDICTORS.....	12
2.4.1 PRIOR ACADEMIC PERFORMANCE.....	13
2.4.2 AGE.....	15
2.4.3 GENDER.....	17
2.4.4 LANGUAGE BACKGROUND.....	18
2.4.5 ENGLISH LANGUAGE PROFICIENCY AS A PREDICTOR.....	20
2.5 CONCLUSION.....	25
CHAPTER 3: PELAS AND VALIDITY.....	27
3.0 INTRODUCTION	27
3.1 POST-ENTRY LANGUAGE ASSESSMENTS (PELAS)	27
3.2 BENEFITS OF IMPLEMENTING PELAS	30
3.3 FEATURES OF PELAS	32
3.3.1 TARGETS: ALL OR SELECTED STUDENTS.....	33
3.3.2 COMPLETION: MANDATORY OR OPTIONAL.....	34
3.3.3 DEVELOPMENT: INTERNAL OR EXTERNAL.....	36
3.3.4 SKILLS ASSESSED: MACRO SKILLS, GRAMMAR AND VOCABULARY KNOWLEDGE.....	37
3.3.5 MODE OF DELIVERY: ONLINE OR PAPER-BASED.....	39
3.3.6 RATERS: HUMANS AS EXPERTS/NON-EXPERTS OR AUTOMATED.....	40
3.3.7 LINKS TO ON-CAMPUS SUPPORT: CONSULTATIONS, WORKSHOPS OR SUBJECTS.....	41
3.3.8 SUMMARY OF PELA FEATURES.....	43
3.4 KEY CONSIDERATIONS OF PELAS	46

3.5 LANGUAGE ASSESSMENT VALIDITY.....	48
3.5.1 BACKGROUND ON VALIDITY AND VALIDATION.....	48
3.6 ARGUMENT-BASED APPROACHES TO VALIDATION.....	50
3.6.1 CRITIQUE OF THE ARGUMENT-BASED APPROACH.....	54
3.7 PELA VALIDATION FRAMEWORK.....	55
3.7.1 EVALUATION INFERENCE.....	57
3.7.2 GENERALISABILITY.....	60
3.7.3 EXPLANATION AND EXTRAPOLATION.....	63
3.7.4 DECISIONS.....	68
3.7.5 CONSEQUENCES.....	71
3.8 CONCLUSION.....	76
CHAPTER 4: METHODOLOGY.....	80
4.0 INTRODUCTION.....	80
4.1 AIMS AND RESEARCH QUESTIONS.....	80
4.2 OVERALL RESEARCH DESIGN.....	82
4.2.1 QUANTITATIVE DATA.....	82
4.2.2 QUALITATIVE DATA.....	83
4.2.3 RESEARCH SITE.....	84
4.3 SAMPLE AND SAMPLING STRATEGY.....	85
4.3.1 STUDENT SAMPLE.....	85
4.3.2 OTHER SAMPLES: ACADEMIC STAFF, BELA USERS AND EXTERNAL RATERS.....	89
4.3.3 SAMPLING STRATEGY AND PROCEDURES.....	93
4.3.4 RECRUITMENT OF PARTICIPANTS.....	93
4.4 ETHICAL CONSIDERATIONS.....	94

4.5 RESEARCH TOOLS.....	95
4.5.1 BOND ENGLISH LANGUAGE ASSESSMENT (BELA).....	95
4.5.2 OTHER TOOLS: BELA 2, MAJOR ESSAY, OVERALL GRADE, & GPA.....	99
4.5.3 QUESTIONNAIRES.....	100
4.5.4 ACADEMIC SKILLS CENTRE FEEDBACK SURVEY.....	103
4.6 DATA ANALYSIS PROCEDURES.....	103
4.6.1 QUANTITATIVE DATA.....	104
4.6.2 QUALITATIVE DATA.....	106
4.7 RESPONDING TO THE RESEARCH QUESTIONS.....	106
4.7.1 EVALUATION INFERENCE.....	106
4.7.2 GENERALISABILITY INFERENCE.....	108
4.7.3 EXPLANATION AND EXTRAPOLATION INFERENCE.....	109
4.7.4 DECISIONS INFERENCE.....	112
4.7.5 CONSEQUENCES INFERENCE.....	114
4.7.6 CAN BELA AND/OR OTHER FACTORS PREDICT ACADEMIC PERFORMANCE OF UNIVERSITY STUDENTS?.....	117
4.8 SUMMARY OF METHODOLOGY.....	117
4.9 CONCLUSION.....	119
CHAPTER 5: RESULTS.....	120
5.0 INTRODUCTION.....	120
5.1 VALIDITY ARGUMENT FOR BELA.....	120
5.1.1 EVALUATION INFERENCE.....	120
5.1.2 GENERALISABILITY INFERENCE.....	137
5.1.3 EXPLANATION AND EXTRAPOLATION INFERENCE.....	142
5.1.4 DECISIONS INFERENCE.....	155

5.1.5 CONSEQUENCES INFERENCE.....	171
5.2 PREDICTING ACADEMIC PERFORMANCE.....	198
5.2.1 PREDICTING ACADEMIC FAILURE.....	199
5.2.2 PREDICTING HIGH ACADEMIC ACHIEVEMENT.....	200
5.2.3 SUMMARY OF RQ2.....	202
5.3 CONCLUSION.....	202
CHAPTER 6: DISCUSSION.....	205
6.0 INTRODUCTION	205
6.1 SUMMARY OF FINDINGS	205
6.1.1 VALIDITY ARGUMENT FOR BELA.....	205
6.1.2 PREDICTING ACADEMIC PERFORMANCE.....	220
6.2 IMPLICATIONS	232
6.2.1 ABILITY TO PREDICT ACADEMIC PERFORMANCE.....	233
6.2.2 THE IMPORTANCE OF THE FIRST YEAR.....	233
6.2.3 WHOLE OF UNIVERSITY APPROACH REQUIRED.....	234
6.2.4 PELAS CAN IDENTIFY AT-RISK STUDENTS.....	235
6.2.5 THE ROLE PLAYED BY ALL UNITS.....	235
6.3 LIMITATIONS AND FUTURE RESEARCH.....	236
6.4 CONCLUSION OF THESIS	239
REFERENCES	241
APPENDICES	273
APPENDIX A: FEATURES OF DIAGNOSTIC ENGLISH LANGUAGE ASSESSMENT (DELA)	274
APPENDIX B: FEATURES OF MEASURING THE ACADEMIC SKILLS OF UNIVERSITY STUDENTS (MASUS).....	277

APPENDIX C: FEATURES OF DIAGNOSTIC ENGLISH LANGUAGE NEEDS ASSESSMENT (DELNA).....	279
APPENDIX D: FEATURES OF POST-ENTRY ASSESSMENT OF ACADEMIC LANGUAGE (PAAL)	281
APPENDIX E: EXTERNAL RATERS.....	283
APPENDIX F: INDICATIVE INTERVIEW QUESTIONS FOR EXTERNAL RATERS..	285
APPENDIX G: BELA VERSION 1	286
APPENDIX G CONTINUED: BELA VERSION 2.....	287
APPENDIX H: ANALYTIC RATING SCALE USED WITH BELA	288
APPENDIX I: RUBRIC FOR MAJOR ESSAY	289
APPENDIX J: STUDENT QUESTIONNAIRE.....	292
APPENDIX K: ACADEMIC STAFF QUESTIONNAIRE	311
APPENDIX L: BELA USERS QUESTIONNAIRE	326
APPENDIX M: SUMMARY OF DATA COLLECTION AND ANALYSIS METHODS .	338

LIST OF TABLES

Table 3.1:	PELA characteristics and key questions	32-33
Table 3.2:	Features of PELA case studies	45
Table 3.3:	Warrants and potential supporting evidence for the Evaluation inference	57-58
Table 3.4:	Warrants and potential supporting evidence for the Generalisability inference	61
Table 3.5:	Warrants and potential supporting evidence for the Explanation and Extrapolation inference	64-65
Table 3.6:	Warrants and potential supporting evidence for the Decisions inference	68-69
Table 3.7:	Warrants and potential supporting evidence for the Consequences inference	72-73
Table 4.1:	Sociodemographic characteristics of student sample	86-87
Table 4.2:	Sociodemographic characteristics of student questionnaire respondents	88-89
Table 4.3:	Roles and number of staff identified as BELA users	90-91
Table 4.4:	Features of Bond English Language Assessment (BELA)	97-98
Table 4.5:	Questionnaire reliability	102
Table 4.6:	Warrants and evidence for the Evaluation inference for BELA	107
Table 4.7:	Warrants and evidence for the Generalisability inference for BELA	108-109
Table 4.8:	Warrants and evidence for the Explanation and Extrapolation inference for BELA	110-111
Table 4.9:	Warrants and evidence for the Decisions inference for BELA	112-113

Table 4.10	Warrants and evidence for the Consequences inference for BELA	114-115
Table 5.1:	Inter-rater reliability per tutor in the original coding and dummy coding ($n = 180$)	127
Table 5.2:	Inter-rater reliability using Fleiss kappa (κ) per external rater 1 in the original coding and dummy coding ($n = 68$)	128
Table 5.3:	Inter-rater reliability using Fleiss kappa (κ) per external rater 2 in the original coding and dummy coding ($n = 52$)	129
Table 5.4:	Test difficulty and test discrimination scores according to each BELA version	136
Table 5.5:	Warrants and supporting evidence for the Evaluation inference for BELA	137
Table 5.6:	Warrants and evidence for the Generalisability inference for BELA	143
Table 5.7:	Warrants and evidence for the Explanation and Extrapolation inference for BELA	155-156
Table 5.8:	Warrants and evidence for the Decisions inference for BELA	171-172
Table 5.9:	Breakdown of scores on BELA 2	188
Table 5.10:	Proportion of students in each BELA category who improved in BELA 2	190
Table 5.11:	Warrants and evidence for the Consequences inference for BELA	197-198
Table 5.12:	Logistic regression analysis of failing after two semesters as a function of demographics and other academic information ($N = 1,115$)	201
Table 5.13:	Logistic regression analysis of obtaining a distinction average after two semesters as a function of demographics and other academic information ($N = 1,115$)	202-203

LIST OF FIGURES

Figure 3.1:	The structure of argument presented by Toulmin (2003)	51
Figure 3.2:	Building blocks of Kane's 1992 interpretive argument	53
Figure 3.3:	Building blocks of the hybrid framework for evaluating PELAs	55
Figure 3.4	Overview of building blocks and inferences contained in the framework for PELA validation	56
Figure 5.1:	Mean comparison between academic staff and BELA users ($N = 25$) of the belief that the criteria used for BELA captured aspects of performance most relevant to academic writing	121
Figure 5.2:	Mean comparison of belief that students completed the task in a quiet environment between students and academics ($N = 111$)	130
Figure 5.3:	Mean comparison of sufficient time limit (60 minutes) for students, academics and BELA users ($N = 130$)	131
Figure 5.4:	Mean comparison of belief that students plagiarise between students and academics ($N = 111$)	132
Figure 5.5:	Mean opinion comparison of the clarity of the task and instructions for both students and academics ($N = 111$)	134
Figure 5.6:	Bar chart illustrating the number of Below satisfactory versus Satisfactory students' scores regarding BELA version ($N = 1,164$)	138
Figure 5.7:	Bar chart illustrating the percentage per sample (pilot versus current study) of BELA results (i.e., Did not complete [DNC], 3, 5, 7 and 9; $N = 1,921$)	140
Figure 5.8:	Bar chart illustrating the percentage per sample (pilot versus current study) number of DNC, Below satisfactory and Satisfactory results ($N = 1,921$)	141
Figure 5.9:	Mean opinion comparison of BELA's similarity with other university tasks for both students and academics ($N = 111$)	144
Figure 5.10:	Mean opinion comparison of whether students brainstormed ideas before starting BELA for both students and academics ($N = 111$)	146

Figure 5.11:	Mean opinion comparison of whether students plan their writing before starting BELA for both students and academics ($N = 111$)	147
Figure 5.12:	Mean opinion comparison between students and academic staff's belief of BELA being representative of students' writing ability ($N = 111$)	148
Figure 5.13:	Mean opinion comparison of BELA as a fair task ($N = 111$)	151
Figure 5.14:	Bar chart illustrating the percentage of unsatisfactory versus satisfactory students' scores regarding their language background (EAL/D; English-speaking background) ($N = 1,164$)	153
Figure 5.15:	Histogram illustrating distribution of Major Essay results for students scoring 3 on BELA ($N = 28$)	156
Figure 5.16:	Histogram illustrating distribution of Major Essay results for students scoring 5 on BELA ($N = 163$)	157
Figure 5.17:	Histogram illustrating distribution of Major Essay results for students scoring 7 on BELA ($N = 339$)	157
Figure 5.18:	Histogram illustrating distribution of Major Essay results for students scoring 9 on BELA ($N = 586$)	158
Figure 5.19:	Mean marks on Major Essay for each BELA score category	159
Figure 5.20:	Histogram illustrating distribution of GPA results after two semesters for students scoring 3 on BELA ($N = 28$)	160
Figure 5.21:	Histogram illustrating distribution of GPA results after two semesters for students scoring 5 on BELA ($N = 163$)	160
Figure 5.22:	Histogram illustrating distribution of GPA results after two semesters for students scoring 7 on BELA ($N = 339$)	161
Figure 5.23:	Histogram illustrating distribution of GPA results after two semesters for students scoring 9 on BELA ($N = 585$)	161
Figure 5.24:	Mean GPA for each BELA score category	162
Figure 5.25:	Mean comparison of sufficient feedback for students between academic staff and BELA users ($N = 27$)	163
Figure 5.26:	Mean comparison of feedback accessibility between students and academics ($N = 107$)	165

Figure 5.27:	Mean comparison of timely feedback for students and BELA users ($N = 118$)	166
Figure 5.28:	Bar chart illustrating the mean of agreement to the meaning of 3 and 5 as below satisfactory ($N = 126$)	168
Figure 5.29:	Bar chart illustrating the mean of agreement to the meaning of 7 and 9 as satisfactory ($N = 126$)	169
Figure 5.30:	Mean opinion comparison of the stigma attached to poor performance for students, academics, and BELA users ($N = 124$)	173
Figure 5.31:	Mean opinion comparison of whether students feel disadvantaged when completing BELA for students, academics, and BELA users ($N = 124$)	175
Figure 5.32:	Mean comparison of perception of BELA as a useful test between students, academics, and BELA users ($N = 123$)	176
Figure 5.33:	Mean comparison of perception of BELA as beneficial between students and BELA users ($N = 116$)	177
Figure 5.34:	Mean comparison of perception of BELA as a waste of students' time between students, academics, and BELA users ($N = 123$)	180
Figure 5.35:	Mean comparison of perception of feedback as useful and informing future learning between students, academics and BELA users ($N = 126$)	181
Figure 5.36:	Mean comparison of perception of SLS (ASC) to improve academic writing between students, academics and BELA users ($N = 60$)	184
Figure 5.37:	Mean comparison of perception that although feedback and support was provided, academic writing did not improve ($N = 43$)	185
Figure 5.38:	Mean comparison of perception of SLS (ASC) and English improvement between students and academics ($N = 43$).	186
Figure 5.39:	Major Essay marks amongst students categorised as having Below satisfactory BELA scores ($N = 191$)	190
Figure 5.40:	Histogram illustrating distribution of Major Essay results in Below satisfactory category who attended ASC ($N = 160$)	191
Figure 5.41:	Histogram illustrating distribution of Major Essay results in Below satisfactory category who did not attend ASC ($N = 31$)	191

Figure 5.42:	Overall grade amongst students categorised as having Below satisfactory BELA scores ($N = 191$)	192
Figure 5.43:	Histogram illustrating distribution of Overall Grades in Below satisfactory category who attended ASC ($N = 160$)	193
Figure 5.44:	Histogram illustrating distribution of Overall Grades in Below satisfactory category who did not attend ASC ($N = 31$)	193
Figure 5.45:	GPA after two semesters amongst students categorised as having Below satisfactory BELA scores ($N = 191$)	194
Figure 5.46:	Histogram illustrating distribution of GPA in Below satisfactory category who attended ASC ($N = 160$)	195
Figure 5.47:	Histogram illustrating distribution of GPA in Below satisfactory category who did not attend ASC ($N = 31$)	195

ABBREVIATIONS

Academic language and learning	ALL
Australian Tertiary Admissions Rank	ATAR
Academic Skills Centre	ASC
Bond English Language Assessment	BELA
Diagnostic English Language Assessment	DELA
Diagnostic English Language Needs Assessment	DELNA
English as an additional language or dialect	EAL/D
English language proficiency	ELP
Grade point average	GPA
International English Language Testing System	IELTS
Measuring the Academic Skills of University Students	MASUS
Post-entry Assessment of Academic Language	PAAL
Post-entry language assessment	PELA
Tertiary Education Quality Standards Agency	TEQSA
Tertiary entrance ranking	TER

CHAPTER 1: INTRODUCTION

1.0 OVERVIEW OF THE STUDY AREA

Globally, increasing access to higher education for all members of society has been an apparent trend (Organisation for Economic Co-operation and Development, 2017), with demand for places at university at peak levels. Higher education is described as being “in a state of massification” (Walker-Gibbs et al., 2019, p. 1) with changing profiles of commencing students due to more inclusive entry criteria and pathway programmes (Grebennikov & Shah, 2012; Murray, 2022; Whiteford et al., 2013). Nonetheless, merely being accepted into a university degree programme does not guarantee a smooth journey through to graduation (Jones, Holder et al., 2000). Improved access and participation in higher education has resulted in increasing numbers of students who commence studies without the required academic literacies to succeed (Lea & Street, 1998). Such literacies include the ability to engage in a range of literacy practices, such as having the requisite academic English language skills to communicate in the discourse of their chosen field (Curnow & Liddicoat, 2008; Lea & Street, 2006).

This thesis titled, ‘Predicting academic performance at university and presenting a validity argument for a post-entry language assessment (PELA)’, was motivated by a desire to comprehend the underlying reasons behind student performance at university in order to maximise student success. It, therefore, set out to investigate the academic performance of university students to determine whether academic and non-academic factors, including results on a post-admission language assessment, predicted academic performance.

This chapter highlights the evolving context of higher education, with a particular focus on the situation in Australia. It presents several challenges faced by higher education institutes and introduces one common response, the implementation of post-entry language assessments known as PELAs. The chapter then details the thesis’s research aims, as well as the significance of the current investigation. Finally, an outline of the subsequent chapters is presented.

1.1 THE EVOLVING CONTEXT AND CHALLENGES WITHIN HIGHER EDUCATION

The change in the student profile, brought about by the demand, is considered an international phenomenon, with higher education growing faster worldwide than what has

been projected (Danilowicz-Gösele et al., 2014; Palmer et al., 2014; Vinke & Jochems, 1993, Webb, 2014). In Australia, higher education has witnessed an evolution in the profiles of commencing students due to changing admission requirements and pathway programmes offered (Burton & Dowling, 2005; Grebennikov & Shah, 2012; Palmer et al., 2014; Smith et al., 2012; Walker-Gibbs et al., 2019; Whiteford et al., 2013). International students made up 28.1% of the total population of students enrolled in Australian universities (Universities Australia, 2022). Domestically, government strategy has reflected a commitment to enhancing access and participation in tertiary education, particularly for populations not traditionally represented (Department of Education, Employment and Workplace Relations, 2009; Whiteford et al., 2013). However, the number of students facing academic difficulties within English medium instruction universities has increased. This has been evidenced by reports in Australian media (e.g., Jackson, 2016) suggesting high numbers of students discontinue their studies (Department of Education and Training, 2018). Concomitant has been concerns surrounding students' lack of requisite academic English language proficiency needed to succeed in their degree programmes (Birrell, 2006; Dodd, 2020; Palmer et al., 2014).

In response to the concerns surrounding language proficiency, the Australian Universities Quality Agency (AUQA) published the Good Practice Principles for English Language Proficiency of International Students in Australian Universities (Department of Employment, Education, and Workplace Relations [DEEWR], 2009), which required institutions to ensure adequate English language proficiency (ELP) of their students from commencement to completion of degree (Humphreys, 2017). Following on from the Good Practice Principles, the Tertiary Education Quality Standards Agency (TEQSA), produced the English Language Standards for Higher Education in 2012, requiring universities to assist in the development of students' ELP as a key graduate attribute for all students at all higher education providers (Baird & Dooney, 2017; Humphreys, 2017; Humphreys & Gribble, 2013).

The English Language Standards for Higher Education presents three common themes. The first theme is the impetus placed on higher education institutions themselves to ensure that students are linguistically prepared for university. This refers to admissions requirements for university, but also includes developing recently admitted students' academic language skills from the onset of their studies. A second theme is the responsibility placed on universities to provide adequate resources in order to assist students to develop

language proficiency throughout their degree. Coupled with this obligation is the required evidence to support the efficacy of support mechanisms put in place. A final theme is the burden of responsibility sitting with students themselves. Students have a responsibility to either maintain or enhance their language skills throughout their degree to ensure that when they graduate, they are proficient in English and can communicate effectively in their chosen field. Overall, the standards demonstrate the requirements of both universities and students to ensure language-related difficulties do not negatively affect academic endeavours (Read, 2019).

1.2 POST-ENTRY LANGUAGE ASSESSMENT

In response to the linguistic challenges and to demonstrate adherence to TEQSA's English Language Standards for Higher Education, mechanisms to assess students' academic language and communication skills early in their studies have been implemented since the 2000s by the majority of universities in Australia. These measures, known as post-entry language assessments or PELAs, aim to identify students who require additional academic language and literacy support to succeed academically. Read (2015b) provided the following definition:

A PELA is administered to students after they have been admitted to a tertiary institution, with a view to identifying those who are likely to struggle to meet the language demands of their degree programme and who should be encouraged or required to enhance their academic language skills. (p. 217)

Therefore, PELAs allow for the assessment of university students' ELP, particularly, academic language proficiency at an early stage in their studies. This, in turn, provides stakeholders with the opportunity to identify deficits in students' communication skills and to implement appropriate support mechanisms.

PELAs were first introduced in Australia in the late 2000s, with most universities currently employing a form of PELA, typically managed by academic language and learning (ALL) units (Read, 2019). Although PELAs are unique to the institution which employs them (Veitch & Johnson, 2022), all aim to enhance or at least contribute to student success at university (Dunworth, 2009; Read, 2015a). Despite their ubiquity in Australia, and their increased prominence internationally (e.g., Fox, 2004; Kim, 2021; Wong et al., 2017), little

empirical validity evidence has been published concerning the assessments. Hence, there is a distinct gap in the literature concerning the lack of validity evidence and research focusing on PELAs as predictors of academic performance. The current research attempted to contribute to this gap by investigating the validity of one PELA, Bond English Language Assessment (BELA) at Bond University, Queensland, Australia.

1.3 RESEARCH OVERVIEW

This thesis set out to achieve two main aims. The first and overarching aim was to determine factors able to predict academic performance amongst an undergraduate student population using available data. This aim addressed growing concerns regarding student attrition at university by attempting to identify characteristics of student performance at one Australian university and the need to support student success. As results on BELA constituted the main measure employed, a second aim was to investigate its validity. It is imperative that any language assessment, such as a PELA, meets the appropriate requirements of quality and results in a meaningful measure of students' abilities (Elder & von Randow, 2008; Read, 2015a). Consequently, an argument for the validity of BELA, encompassing both theory and empirical evidence, was required so that stakeholders (e.g., students, academic staff and administrators) can make informed decisions about enhancing students' language skills. An established PELA validity framework developed by Knoch and Elder (2013) was employed to present the validity argument for BELA.

In order to achieve these aims, two research questions were investigated. The first research question posed was whether a satisfactory validity argument could be presented for the BELA. The second research question was whether BELA and/or other factors were able to predict academic performance of university students.

1.4 SIGNIFICANCE OF THE CURRENT INVESTIGATION

If students are not equipped with adequate language skills to participate in higher education, they will likely not succeed at university, discontinue their studies, and potentially be lost to the university system. These students may become "failed models" and, subsequently, discourage members of their community from pursuing higher education. The current investigation makes a significant contribution to the literature concerning English medium instruction universities by developing a profile of students who may require additional academic support. Such a profile may help identify students potentially at risk of

academic failure, which in turn allows decision makers to ensure the provision of appropriate support and, subsequently, enhance student retention rates. Generalisable findings from this study may be utilised by other tertiary institutions. Hence, this research contributes to best practices regarding identification of students at risk and, through early assistance, improving overall student tertiary experience and positively affecting retention and success rates.

1.5 OUTLINE OF CHAPTERS

This thesis comprises seven chapters. Chapter 1 presented the overview of the current study area and challenges faced within the context of higher education. It then introduced commonplace measures of language ability used post-admission known as PELAs. The overview, including aims and research questions of the current study, were presented. Finally, the significance and contribution of the study were detailed.

Subsequently, Chapter 2 focuses on academic performance predictions. It begins by reviewing the literature regarding current challenges faced by universities in terms of student attrition, retention and success. It then examines the implementation of academic performance predictions in response to such challenges. Predictive factors are reviewed to ascertain commonly cited variables. Language proficiency or ELP as a predictor is then analysed, with a focus on IELTS due to the substantial body of work on the assessment and, finally, PELAs as predictors of academic performance.

Chapter 3 presents a literature review of PELAs, including a definition, justifications for their use, characteristics, and key considerations. The most salient consideration, validity, is then reviewed from a theoretical lens. Finally, the chapter examines the PELA validity framework.

Chapter 4 presents the methodology of the thesis. In the chapter, the research site is presented, along with the overall research design, sample and sampling strategy undertaken, ethical considerations, research tools, and data analysis procedures. The main measure, BELA, is analysed. The chapter concludes with a presentation of how the research questions were answered.

In Chapter 5, the results, including quantitative and qualitative data analyses, are presented. The final chapter of the thesis is Chapter 6, which contains the discussion of the research findings. It responds to the main aims of the study and presents answers to the

research questions, including implications, limitations and opportunities for future research.
The chapter closes with a conclusion to the thesis.

CHAPTER 2: PREDICTING ACADEMIC PERFORMANCE

2.0 INTRODUCTION

The main aim of this thesis was to explore factors that were able to predict the academic performance of undergraduate students. This chapter, therefore, reviews the literature on academic performance at university. Firstly, the chapter presents background on academic performance, including a definition. It then analyses the overarching challenge universities face in terms of attrition, retention and success, leading to a rationale for conducting academic performance predictions. The chapter subsequently presents an overview on predictive studies in academia and then analyses previous academic performance as a predictive factor, followed by non-academic factors, including age, gender, and language background. Finally, the chapter outlines the relationship between English language proficiency or ELP and academic performance, focusing on a high stakes language test, for which there is an abundance of empirical studies, and then PELAs.

2.1 BACKGROUND ON ACADEMIC PERFORMANCE

This chapter first defines academic performance in the context of this thesis. Academic performance takes into consideration students' results in relation to several factors, such as individual subjects, multiple semesters and length of study. Student results can refer to their performance in one semester of study or several and may include the results of students who have graduated, illustrating a longitudinal picture of academic performance. The most commonly utilised and referred to results in predictive studies have proven to be students' grade point average or GPA (Belingham, 1993; Bridgeman et al., 2016; Hill et al., 1999; Humphreys et al., 2012; Li, 2015; York et al., 2015). GPA is defined as "the overall averaged subject-matter grade point" (Strickland, 1998, p. 3). It includes semester GPA, which is the average of all grades awarded in one semester and cumulative GPA, which is the average of all grades awarded in all of the subjects attempted towards completion of a degree programme. Other means of determining academic achievement include self-reported data, weighted average marks (WAM; Harvey & Luckman, 2014; Knoch et al., 2016), which focuses on the average marks obtained in subjects as percentages rather than grades (e.g., 75% cf Distinction), success rate (or "pass rate analysis"; Elder, Bright et al., 2007), which is the proportion of subjects that have been successfully completed, and its opposite, failure rate (Scouller et al., 2008). Different studies have adopted different measures; however, GPA

stands out as the most commonly utilised measure of academic performance (Ihlenfeldt & Rios, 2022; York et al., 2015). For the purposes of this thesis, GPA was utilised as the measure of students' academic performance due to it being typically referred to in predictive research and utilised at the research site.

2.2 THE PROBLEM OF ATTRITION, RETENTION AND SUCCESS

Having presented a background of academic performance, it is now important to identify the overarching problem this thesis aims to address. Globally, increases in the number of enrolled domestic and international university students have not resulted in improved student outcomes (Anderton et al., 2016). Unfortunately, due to a plethora of reasons, attrition rates are higher than desired; retention and success rates also fall below ideal scenarios. It is, therefore, necessary to examine these three inter-related terms.

The first term to be considered, attrition, is defined as the percentage of students who neither continue studying at a particular university nor graduate from their degree programme (Grebennikov & Shah, 2012; Tertiary Collection of Student Information, n.d.a). Attrition continues to be a pervasive issue faced by higher education institutions globally (Brunsden et al., 2000; Grebennikov & Shah, 2012). In Australia, 2020 data demonstrated a 12.44% attrition rate amongst all undergraduate university students, was (Department of Education, 2023b). For students enrolled in a sub-bachelor programme, such as a diploma, the attrition rate increased to 31.01% in the same year (Department of Education, 2023b). Attrition, problematic at several levels, carries with it various negative outcomes faced by students, institutions and society more broadly. At an individual level, attrition wastes students' time and effort and results in a lack of enrichment for them, which can have serious impacts on self-esteem and increased financial and opportunity costs (Cao & Gabb, 2006; Smith et al., 2012; Thomas et al., 2021). For tertiary institutions, attrition can tarnish the reputation of universities impacting the ability to attract new students (Ozga & Sukhnandan, 1998) and harm the sector as a whole. Attrition can significantly impact institutions financially, through loss of fees, increased recruitment and tuition costs, and loss of potential government funding. As illustrated by Grebennikov and Shah (2012, p. 224), "for each onshore international student failing to return after year one of a three year undergraduate program one large metropolitan public university loses approximately \$23,000 [AUD]." Moreover, Adams et al. (2010) suggested domestic student attrition costs Australian higher education \$1.2 billion AUD per year. At a societal level, high rates of attrition impact the intellectual

capacity of the nation's human capital, which includes sufficient skilled labour to keep up with the evolving needs of industries and contributing to the wealth and well-being of society (Grebennikov & Shah, 2012). Although a level of student attrition is unavoidable (Ozga & Sukhnandan, 1998), it has far-reaching impacts. Consequently, tertiary institutions must work to understand characteristics of students who discontinued their studies including their unique situations to provide a supportive and nurturing environment for them to flourish (Harvey & Luckman, 2014; Long et al., 2006).

The next concept, retention, in contrast to attrition, refers to the percentage of students who continue their studies after the first year (Tertiary Collection of student Information, n.d.b). In Australia, 2020 data depicted a retention rate of 87.38% amongst undergraduate university students (Department of Education, 2023b). In comparison, the retention rate for students enrolled in sub-bachelor programmes dropped to 67.56%. Retention of students at tertiary level has been “one of the high-priority government targets in developed countries such as Australia, the UK and the USA” (Grebennikov & Shah, 2012, p. 234); therefore, it has been researched extensively. The literature describes enhancing retention as a multifactorial challenge, unable to be overcome via a single solution (Kift & Nelson, 2005; Lobo, 2012; Trotter & Roberts, 2006). It is consistently argued that intervention, if necessary, needs to occur as early as possible in students' academic endeavours to assist them in achieving their goals (Nelson et al., 2012; TEQSA, 2020). As commencing students can struggle with the initial requirements of academia (Bright & von Randow, 2008; Smith et al., 2012; Tynan & Johns, 2015), interventions in the first year of study enhance retention (Tumen et al., 2008). Consequently, the first-year experience is considered to be “critical in terms of retention, satisfaction and graduate outcomes” (Smith et al., 2012, p. 91), and students' academic achievements after two semesters have been shown to be indicative of their performance across the entirety of their study programmes (Gershenfeld et al., 2015; Humphreys et al., 2012). Numerous studies have demonstrated the importance of early intervention (Fox, 2004; Roessingh & Kover, 2002; Watt & Roessingh, 2001), which may involve, for instance, identifying students potentially at risk of not coping academically to offer support, to increase their academic confidence (Scouller et al., 2008).

A third and final term requiring analysis is success, as it too has become “one of the most widely used constructs in educational research and assessment within higher education” (York et al., 2015, p. 1). Academic success presents as a highly complex construct with

varied definitions (Lydster & Murray, 2018; York et al., 2015), including academic achievement. One measure of success is success rate, described by the Tertiary Collection of Student Information (n.d.c) as the equivalent full-time study load of subjects passed in comparison with the number of subjects attempted. In Australia, the success rate of undergraduate university students in 2021 was 85.58% (Department of Education, 2023b), whilst for sub-bachelor programme students, this figure was 76.70%. In addition to success rate, grade point average or GPA is again commonly utilised (York et al., 2015) to measure student success. Similar to success rate, GPA indicates whether students are successfully completing their subjects and provides an indicator of just how well, or poorly, students are performing at university. In the context of academia, success is typically defined as successful completion of subjects (i.e., passing), which is purely academic success based on students passing subjects they attempt, without consideration of additional factors, such as confidence, sense of achievement and employability. Students passing their subjects is desirable for all stakeholders; therefore, understanding the factors that may play important roles in this regard is critical.

Considering the significance of attrition, retention and success, efforts to maximise university student success at all levels have been introduced. At a government level in Australia, these efforts have included two significant changes. First, the Higher Education Support Amendment (Job-Ready Graduates and Supporting Regional and Remote Students) Bill 2020 No. 93, 2020 to the Higher Education Support Act 2003, No. 149 was legislated. Under the legislation, students with low completion rates, defined as a fail rate of higher than 50%, would lose access to Commonwealth funding (StudyAssist, n.d.). The amendment was criticised (e.g., Norton, 2022; Warburton, 2021) and later scrapped in 2023 (Department of Education, 2023a). Second, the Australian Universities Accord was commissioned in November 2022. The Accord aims to improve the accessibility, affordability, quality and sustainability of higher education in Australia (Department of Education, 2023a). To achieve these aims, recommendations and performance targets will be established to achieve long-term security and prosperity for the sector and the nation (Department of Education, 2023a). The Accord is in its infancy; however, its focus on equity of access and changes to university funding suggest that attrition, retention and success will be high priorities for universities.

2.3 PREDICTING PERFORMANCE IN RESPONSE TO THE PROBLEM

Following research into attrition, retention, success, and changes to legislation, priority has been placed on supporting the academic performance of university students (Scouller et al., 2008). Students who may experience academic difficulties, particularly early in their studies, must be noticed by stakeholders within universities who can assist. One strategy in response to this priority involves predicting which students may require additional support in order to succeed academically. Such systems of detecting students potentially at risk of failure early in their studies are important in mitigating attrition and contributing to retention and success. As argued by Harvey and Luckman (2014, p. 21), “central to preventing attrition is the ability to predict attrition.”

This section identifies several beneficial outcomes of conducting such studies for institutions and, most importantly, for students themselves. For institutions, predictive studies create opportunities to increase retention and success rates by identifying students who are at risk of being unsuccessful and, thereafter, discontinuing their studies (Smith et al., 2012). Predictive studies allow effective and efficient determination of students who may require additional support to ensure they reach their academic potential. Pinpointing the characteristics of students who have succeeded and of those who have not succeeded at university allows for an understanding of the unique situations in-coming students are presented with.

Understanding the attributes that may contribute to students being at risk of potential academic failure allows universities to establish appropriate and effective support mechanisms. Predictions, therefore, assist in the provision of a supportive and nurturing environment in which commencing students can succeed. This, in turn, allows for the equitable provision of support in order to maximise student outcomes. Academic performance predictors, encompassing both success and failure, are important endeavours for support service provision. As noted by Elder (1993, p. 88), predictive studies “have an important part to play in minimizing wastage within educational systems.”

Moreover, predicting academic performance may assist educators and course convenors in adapting teaching and learning environments to create smooth transitions from secondary to tertiary education (Burton & Dowling, 2005). Predictive studies assist in curriculum enhancement, which benefits all students and facilitates a move away from the

dated focus on “the student as the problem” (Ozga & Sukhnandan, 1998, p. 316). Consequently, predictive studies benefit entire cohorts of students via enhancement of course curricula (Anderton et al., 2016). Such enhancements lead to positive learning experiences and, ideally, increased success and graduate outcomes concerning future employment opportunities (Affendey et al., 2010). Having established the benefits of predictive studies, an overview of the commonly cited academic performance predictors can now be presented.

2.4 OVERVIEW OF COMMONLY CITED PERFORMANCE PREDICTORS

Literature on predicting academic performance demonstrates a plethora of academic and non-academic variables that play a role in students’ performance at university. Academic factors include performance in previous studies, such as high school or other university studies (Scott & Smart, 2005). On the other hand, non-academic factors are typically categorised into demographic and psychosocial factors (Grebennikov & Shah, 2012). These include age, gender, language background, language proficiency or ELP and academic literacies (van Dijk, 2015). Other factors include socio-economic status (Walker-Gibbs et al., 2019) and psychological factors, such as personality factors, self-efficacy and motivation (Burton & Dowling, 2005; Graham, 1987; Ho & Spinks, 1985; Malczewska-Webb, 2016; Mthimunye & Daniels, 2019), learning styles and study skills (McKenzie & Schweitzer, 2001), lack of information about the degree and possible the career pathways available (Grebennikov & Skaines, 2009), chosen field of study and compatibility, involvement in campus life and integration into the university (Grebennikov & Skaines, 2009; McKenzie & Schweitzer, 2001), financial concerns (Lobo, 2012), and level of employment (Cao & Gabb, 2006; McMillan, 2005).

The literature has identified many factors contributing to students’ academic performance (Bayliss & Ingram, 2006; Bridgeman et al., 2016; Feast, 2002), highlighting the need to examine factors central to academic performance. Notably, no single set of factors that can predict academic performance have been determined (Smith et al., 2012). The next sections of this chapter present the most commonly referred to academic and non-academic factors and examines findings regarding academic performance predictions. The factors identified include previous academic performance, age, gender, language background, and ELP.

2.4.1 PRIOR ACADEMIC PERFORMANCE

First and foremost, students' previous academic performance has proven to be a "powerful" (Tumen et al., 2008, p. 234) factor when predicting performance at tertiary level. Generally, the literature supports the view that students who commence university studies with higher entry scores via secondary school achieve higher results (Johnes, 1997). Furthermore, previous experience at university has been found to predict academic success (Asif et al., 2015; Ruegg et al., 2020). As previous academic performance, whether it be in high school, in previous tertiary studies, or at early stages of tertiary studies, is recognised as a powerful predictor of academic performance, numerous studies have investigated the relationship.

Performance in secondary school has been demonstrated to be a sound predictor of academic performance at tertiary level in both Australian and international education institutions. In the Australian context, previous performance in high school is commonly measured by tertiary entrance rank or TER. Numerous studies have found a significant correlation between TER scores and academic performance (Anderton et al., 2016; Dobson & Sharma, 1993; Evans & Farley, 1998; McKenzie & Schweitzer, 2001). A study of 197 first semester Faculty of Science ($n = 149$) and Faculty of Information Technology ($n = 48$) students at Queensland University of Technology by McKenzie and Schweitzer (2001), for example, found prior academic performance to be the most significant predictor of academic performance. Furthermore, Burton and Dowling's (2005) study indicated that the Overall Performance (Queensland's previously used equivalent to TER) scores of 132 high school graduates significantly predicted academic performance amongst on-campus, engineering students at the University of Southern Queensland. This finding was consistent with a much larger study by McKenzie et al. (2004) of first-year engineering students ($N = 1,193$) at Queensland University of Technology.

Globally, research outcomes demonstrate similar findings. For example, a large German study by Danilowicz-Gösele et al. (2014), which examined the performances of 12,315 students at a German university, determined a significant, strong relationship between performance in secondary school subjects and academic success (i.e., graduation/completion) at university. The researchers stated that "the high school leaving grade is by far the best predictor of both the probability of graduating and the final grade obtained at university" (Danilowicz-Gösele et al., 2014, p. 21). Similarly, a systematic review of literature between

1997 and 2010 and comprising 7,167 studies and 241 data sets in the US by Richardson et al. (2012) found high school GPA, as well as Scholastic Aptitude Test (SAT) scores, had medium sized correlations with tertiary level GPA.

However, this viewpoint has been refuted by other studies, which have contradicted the aforementioned findings. For example, Dobson and Skuja (2005) found that although high entrance rankings were sound predictors of tertiary performance, this was limited to certain disciplines. Student performance in disciplines such as Information Technology, business, humanities and creative arts was found to be inadequately predicted by high school performance, whilst health and education degrees were said to be poorly predicted by previous results (Dobson & Skuja, 2005). Furthermore, the works of Jones, Holder et al. (2000) and Holder et al. (1999) determined that university entrance scores from domestic, secondary school pathways, when utilised solely, were not good predictors of academic performance amongst Pharmacy degree students. Overall, research outcomes have concluded that prior academic performance alone was insufficient as a predictor.

Turning to performance at early stages of university, research has found a significant relationship with academic performance later in students' degrees. Students who have completed one semester or one year of study, for instance, with higher GPAs are more inclined to perform better in the remainder of their studies (Gershenfeld et al., 2015). Conversely, students with low GPAs continue this trend and, in many instances, discontinue their studies. For instance, a longitudinal study of 7,314 undergraduate students by Tumen et al. (2008) at the University of Auckland found that students' progress in their degrees and their GPAs were two important predictors of different academic outcomes amongst students. Similarly, Cao and Gabb (2006) argued the most powerful predictor of attrition amongst students at Victoria University was low levels of academic achievement. With a sample of approximately 12,000 students commencing their studies over three years, this finding highlighted the necessity for early identification of students who do not perform satisfactorily in their first semester and, more importantly, the need for timely intervention. Support was determined by Grebennikov and Skaines (2009), who found that students with low GPA were more inclined to leave the University of Western Sydney, without applying to study elsewhere, indicating they were lost to the tertiary system. GPA was found to be related to language background (in this case, EAL/D status), gender (i.e., male), and low socio-economic status among others. Furthermore, in a US study, Gershenfel et al. (2015) found

that first semester GPA, as opposed to cumulative GPA after years of study, was “a strong early predictor” (p. 483). Those achieving a first semester GPA of 2.33 (out of 4.0) or below were at significant risk of not completing their studies. Specifically, an odds ratio of .47 was determined, indicating these students were almost half as likely to graduate compared to those maintaining a GPA of above 2.33. Finally, Harvey and Luckman (2014) found that performance in the first year of an undergraduate Arts degree at LaTrobe University, Victoria was the strongest predictor of student attrition ($N = 1,124$). As opposed to GPA, the researchers used WAM and Success Rate and found “a relatively strong relationship ... between poor marks/failure rates and attrition” (Harvey & Luckman, 2014, p. 25) with highly significant correlations of 0.45 and 0.44 for WAM and Success Rate, respectively. This study is notable as the majority of Australian predictive studies have focused on TER and attrition/retention, with a relative dearth of literature focusing on the relationship between university marks and attrition/retention. In summary, research outcomes recognise that students’ academic performance at the commencement of their studies can predict performance later on.

In summary, previous academic performance has proven to be a factor when predicting performance at university. Research findings have perceived academic performance at secondary school as a sound predictor of academic performance at tertiary level in both Australian and international studies. However, there are contrasting findings, with studies revealing that prior academic performance alone was inadequate as a predictor of university performance. Although inconsistencies in the literature are apparent, the plethora of research that has found a correlation between prior academic achievement in secondary and the first year of tertiary education provides further justification for developing a “warning system” of students who are potentially at-risk of withdrawal (Harvey & Luckman, 2014, p. 26). Additionally, performance at early stages of university studies has been found to have a positive relationship with academic performance later on in the degree programme. Students with low GPAs early in their studies are inclined to continue this trend and/or discontinue studying. Again, universities have an opportunity to create a system of early detection of poor performance amongst their cohorts.

2.4.2 AGE

A further commonly investigated factor in predictive studies is age, for which mixed findings have been determined with some studies finding younger students had more positive

outcomes than their older peers, while other determined the opposite. As noted by McKenzie and Schweitzer (2001), some studies have found younger students (e.g., recent school leavers) attain higher results and are more inclined to persist with their studies in contrast to more mature students. Specifically, students 19 years and younger have been found to have higher completion rates as compared to other age groups (Cao & Gabb, 2006; Department of Education, Skills and Employment, 2021). For instance, Shah and Burke (1996), an Australian study, found that a 20-year-old undergraduate student has the highest probability of degree completion. Cao and Gabb's (2006) study examining 12,500 commencing students and their attributes at Victoria University in 2002 and then again in 2004 found that older students had higher attrition rates than those in their early 20s. In addition, Vinke and Jocham's (1993) research involving 90 postgraduate Indonesian engineering students at a Dutch English medium of instruction university, found younger students to be more likely to successfully complete their degree programme compared to mature students. The researchers opined that mature students have greater difficulty adapting to a different academic culture. The research supports the view that younger students achieve more positive academic outcomes compared to more mature students.

In contrast, other studies investigating age as a predictor of academic performance demonstrate an opposing view, which highlight inconsistent findings (Anderton et al., 2016; McKenzie & Schweitzer, 2001; Johnes, 1997). A study of 325 students, which aimed to produce a predictive model identifying students at risk in a first-year accountancy subject at Edith Cowen University, Perth, determined age was a contributing factor (Smith et al., 2012) with younger students (i.e., below 25 years) identified as more at risk. Similarly, a New Zealand study by Scott and Smart (2005) found that when many other variables (e.g., sex, ethnicity, previous qualification) were controlled for, older students were more likely to complete their degrees than their younger peers. This finding was later supported by another longitudinal New Zealand study by Tumen et al. (2008), which again found that mature students (i.e., aged 20 years and above) were more likely to complete their degrees faster than recent high school graduates. Overall, age was found to exhibit "a minor effect on the likelihood of departure from a program" (Tumen et al., 2008, p. 248) upon controlling for student achievement and factors related to study. A systematic review by Mthimunye and Daniels (2019), which investigated predictors of performance of undergraduate nursing degree students, also found that the older a student was, the higher the chances they would achieve desired results.

In sum, there are mixed findings in the literature concerning the relationship between age and academic performance. Numerous studies have found that younger students (e.g., recent school leavers aged between 17 and 20 years old) attain higher results and were more inclined to persist with their studies compared to mature students. In contrast, other studies' determined that younger students (i.e., younger than 25 years old) are more at risk of academic failure, and older students are more likely to complete their degrees in comparison to recent high school graduates. Age in itself is stated not to be a factor affecting the academic achievement of university students; however, it is related to other factors, such as prior content knowledge, which may immediately affect success at tertiary level (Vinke & Jocham, 1993).

2.4.3 GENDER

Similar to age, numerous studies analysed and compared educational performance of males and females with mixed results. Some studies have found that females are more likely to complete their degree programmes compared to their male counterparts and demonstrate higher academic achievement compared to males. In comparison, research outcomes have found males perform worse than their female counterparts. Finally, several studies have found that when gender is identified as a significant factor, the margin in terms of performance is minimal.

Studies have revealed that females are more likely to complete their degree programmes compared to their male counterparts. Females also tend to demonstrate higher academic achievement compared to males (Cao & Gabb 2006; Grebennikov & Skaines, 2009; Mthimunya & Daniels, 2019). Anderton et al. (2016) investigated which variables predicted grades on a health sciences subject. Significant factors included gender, with females performing 2.8% to 7.8% better than males over a three-year investigation. This was statistically significant in two of the three years (i.e., 2012 and 2014, not 2013). In comparison, a study of 8,896 undergraduate students at the University of Western Sydney found an association between gender and low GPA (Grebennikov & Skaines, 2009). Specifically, the researchers found males to maintain a low GPA in contrast to females. Similarly, Smith et al.'s (2012) study determined males were more at risk of failure in a first-year accounting subject than females. Overall, research reveals females achieve greater academic success compared to males.

A number of studies, however, have found that if and when gender is a significant factor, the margin between the genders is minimal. Cao and Gabb's (2006) study examining 12,500 commencing students and their attributes at Victoria University in 2002 and then again in 2004 found that males only had a slightly higher rate of attrition. Consistent with this finding was Danilowicz-Gösele et al.'s (2014, p. 21) research which found gender and other factors "play only a minor role." Other studies have found there was no performance difference between females and males in science (MacKenzie & Schweitzer, 2001) or medicine for graduate students (Puddey & Mercer, 2014). Additionally, when controlling for student experiences and achievement, Tumen et al. (2008, p. 245) found that gender, as well as ethnicity and socio-economic status, had "no predictive power". Harvey and Luckman (2014) also found that demographic factors, such as gender, age, socio-economic status, and regionality, had no significant relationship with attrition within a sample of 1,124 commencing undergraduate Arts students at LaTrobe University, Victoria.

In sum, a plethora of studies have investigated educational performance of males and females. Most studies have found that females are slightly more likely to complete their degree programmes and demonstrate higher academic achievement compared to males. A number of studies, however, have found minimal to no difference between the sexes regarding performance at university. When differences were found, the margin was usually minimal, with males having only a slightly higher rate of attrition. These findings suggest that gender as a single factor is unreliable to predict academic performance.

2.4.4 LANGUAGE BACKGROUND

Another factor commonly investigated amongst applied linguistic literature is the relationship between language background and academic performance. Overall, there are inconsistent findings. Some studies have found that students for whom English is an additional language or dialect (EAL/D) perform more poorly than students from English-speaking backgrounds. However, other studies have revealed that EAL/D students perform better academically than their English-speaking background peers.

Research outcomes have indicated that EAL/D students perform below their English-speaking background counterparts. This has been found regarding nursing (Bougnan, 1993; Gerardi, 1996) and accountancy (Smith et al., 2012). For example, in a first-year accountancy subject at Edith Cowen University, Perth, language background was the most important

variable in distinguishing between students who succeeded and students who failed the subject (Smith et al., 2012). Overall, 68% of English-speaking background students passed the subject in contrast to 48% of EAL/D students. Furthermore, in a study of undergraduate students at the University of Western Sydney, EAL/D status was associated with low GPA (Grebennikov & Skaines, 2009). EAL/D students had lower GPAs than their English-speaking background peers.

In contrast, other studies have refuted the view that EAL/D students perform below English-speaking background students. A Canadian study by Berman and Cheng (2010) examined perceived difficulty of various language skills amongst EAL/D and English-speaking background students, as well as academic outcomes. The study found EAL/D students perceived language skills to be more difficult in their studies compared to English-speaking background students. However, in terms of actual performance, the study revealed that undergraduate EAL/D students achieved similar results (Berman & Cheng, 2010). Studies conducted at the research site displayed similar findings (Webb, 2013, 2014, 2015), with Malczewska-Webb (2016) positing a self-efficacy as a key factor in EAL/D students' success. Other studies have demonstrated that language status was unrelated to GPA (e.g., Wu Pong et al., 1997). The study by Cao and Gabb (2006), examining 12,500 commencing students and their characteristics at Victoria University found similar attrition rates for both EAL/D and English-speaking background students. Notably, a study by McMillan (2005) determined that EAL/D students, in fact, had lower attrition rates than English-speaking background students, thus performed better academically than their English-speaking background counterparts. These studies reveal that language background may not play a role in academic performance.

Language background or language status, whether it be English speaking-background or EAL/D, has been identified as a significant factor in academic performance predictions; however, inconsistent findings appear. On one hand, EAL/D students have performed more poorly than their English-speaking background student peers, particularly in accountancy, nursing and other undergraduate programmes. On the other hand, studies have found minimal to no difference between academic performance of EAL/D and English-speaking background students. EAL/D students performed better academically in comparison in one investigation. It appears that discipline and programme level (i.e., undergraduate or postgraduate) play a

role in whether language background predicts academic performance. Therefore, more contextualised research is required.

2.4.5 ENGLISH LANGUAGE PROFICIENCY AS A PREDICTOR

In the fields of applied linguistics and language assessment, a commonly researched predictor of academic performance is students' English language proficiency or ELP. Predictive studies examining the correlation between ELP and subsequent performance at university date back to the 1960s with research including the US studies by Mulligan (1966) and Sugimoto (1966), which focused on academic success and completion of programme respectively. Both Mulligan (1966) and Sugimoto (1966) found a minimal relationship between ELP and academic performance. Since the studies of Mulligan (1966) and Sugimoto (1966), a growing body of literature has demonstrated an important, yet uncertain link between ELP and performance at university (Daller & Phelan, 2013; Feast, 2002; Scouller et al., 2008). Overall, research points to an association between ELP and academic performance; however, studies have found that there may be a minimal to weak relationship. This section presents an initial overview of the literature concerning ELP and academic performance. Subsequently, it examines findings concerning research into one of the most prevalent tests of ELP, International English Language Testing System (IELTS). The section then focuses on more nascent measures, post-entry language assessments or PELAs.

On the one hand, research outcomes point to the important relationship between ELP and performance at university. Adequate ELP has been identified as a necessary factor in terms of success within English-medium instruction tertiary institutions (Bridgeman et al., 2016; Fox, 2004; Mthimunye & Daniels, 2019). Daller and Phelan (2013, p. 188), for instance, argued that ELP is "a crucial factor" for academic success amongst students, irrespective of their chosen field. Similarly, Scouller et al. (2008, p. 176) described adequate language proficiency for all students as "extremely important" for university success. Empirical studies have demonstrated that at low levels, language proficiency commands increased importance as a variable when predicting academic performance (Elder, 1993; Feast, 2002). In support, other studies have found that inadequate ELP seriously hinders academic performance and success of students (Feast, 2002; Knoch & Elder, 2013; Lloyd-Jones et al., 2011).

On the other hand, scholars have argued against the existence of a strong relationship between ELP and performance at university, due to other factors influencing academic achievement (Bayliss & Ingram, 2006; Bridgeman et al., 2016; Feast, 2002). These factors include several academic and non-academic factors as discussed earlier, such as prior academic performance (Bellingham, 1993; Bridgeman et al., 2016). It is important to note that being highly proficient in English does not ensure high academic performance (Daller & Wang, 2017; Heeren et al., 2021; Martirosyan et al., 2015; Riazi, 2013). As argued by Humphreys (2022, pp. 3-4), “language is a key enabler (though not a guarantee) of academic success.” As a result, the precise relationship between ELP and academic performance at university remains unclear. In 1987, Graham (p. 513) concluded that examining the association between ELP and academic performance “does not reveal clear-cut answers”, describing the relationship as “murky indeed”. Almost four decades later, inconsistencies and mixed findings are apparent (Elder, 2017; Pearson, 2021). Consequently, it is important to the review studies investigating the relationship between ELP, as measured by a high stakes, commercial assessment (i.e., IELTS), and academic performance.

Much of the literature utilises the International English Language Testing System or IELTS, considered one of the most influential and prevalent tests of ELP (Ihlenfeldt & Rios, 2022; Mauriyat, 2021; Pearson, 2021). Several studies have determined a strong to moderate relationship between relationship between IELTS and GPA (Bellingham, 1993; Humphreys et al., 2012; Yen & Kuzma, 2009). For instance, Woodrow (2006) determined a significant association between IELTS and GPA. The variance ranged between 16% for overall IELTS scores and 22% for the listening subtest score, meaning 16% and 22% of students’ GPA was explained by their overall IELTS and listening scores respectively. Hill et al. (1999) found a significant, moderately strong correlation between overall IELTS scores and GPA for students, equating to an uncommonly high 29% variance in GPA. In comparison, studies have also found weak, positive correlations between IELTS scores and academic performance (Avdi, 2012; Cotton & Conrow, 1998; Elder, 1993; Feast, 2002; Kerstjens & Nery, 2000). A meta-analysis by Ihlenfeldt and Rios (2022), which synthesised the results of 32 studies comprising 132 effect sizes between 2006 and 2021, found IELTS exhibited a significant yet weak, positive correlation with GPA. Notably, much of the literature points to lower language proficiency having the strongest impact on academic performance. In contrast, other studies have found no relationship between IELTS and GPA (Craven, 2012; Ingram & Bayliss, 2007). Dooney’s (1999) study found no correlation between speaking,

writing and listening scores and GPA; a significant, albeit weak correlation between reading scores and GPA (0.34), accounting for 12% of students' GPA, were determined. Dooley and Oliver (2002, p. 36) later concluded, "little evidence [exists] for the validity of IELTS as a predictor of academic success." In summary, most studies confirm a relationship between IELTS and academic performance. However, it is important to acknowledge that inconsistencies in the literature are apparent.

The inconsistent findings regarding IELTS as a predictor of academic performance are primarily due to the variety of research designs and methods employed. In a systematic review of 32 studies from 1989 to 2019 comprising 5,426 students, Pearson (2021) confirmed that despite many studies being correlational, there were wide ranging research designs employed, including the use of self-report data, data collection, analysis, interpretation and reporting. Furthermore, inconsistencies appear due to the inclusion of samples within one or two degree programmes at a single institution (Bridgeman et al., 2016). Having small sample sizes restricts the range of scores on the test (Bridgeman et al., 2016; Graham, 1987), resulting in "a truncated sample" (Alderson et al., 1995; Daller & Phelan, 2013; Fox, 2004; Ginther & Yan, 2018; Kerstjens & Nery, 2000). A final reason for mixed findings is that predictive studies rarely include students scoring high on language assessments. Not including students with high scores on pre-entry assessments results in "a depressed validity coefficient" (Ferguson & White, 1998, p. 16). Overall, examining the relationship between IELTS and academic performance has revealed that despite a lack of consensus, research suggests that IELTS scores can be used to predict academic performance. Inconsistencies are apparent due to the variety of research designs and methods employed throughout the literature. This section now examines the relationship between PELAs and academic performance.

Compared to the plethora of research using IELTS as a predictor of academic performance, there is a paucity of studies investigating PELAs' role. Four Australasian university PELAs have been the focus of prediction studies: Diagnostic English Language Needs Assessment (DELNA), Measuring the Academic Skills of University Students (MASUS), Notre Dame University's PELA, and James Cook University Singapore's PELA. Research outcomes have revealed a weak to moderate association between PELAs and students' academic performance.

The most widely researched PELA regarding predictive studies has proven to be the University of Auckland's Diagnostic English Language Needs Assessment (DELNA). Such investigations have occurred since its inception with the development and validation final report noting, "a moderate relationship could be taken as evidence that language is a significant factor in student performance at the university" (Elder & Erlam, 2001, p. 22). Since the validation report, comprehensive studies of DELNA have found a moderate relationship between results on the PELA and academic performance (Davies & Elder, 2005; Fox, Haggerty et al., 2016). Typically, research demonstrates variances of approximately 10%, meaning scores on the PELA explain 10% of student academic performance. Elder, Bright et al. (2007), utilising correlational and pass rate analysis with samples of 761 and 1,052 over two semesters, determined a significant, yet weak association between language proficiency (i.e., DELNA scores) and subsequent first semester GPA. The PELA results explained between 5% to 17% of students' academic performance (Elder, Bright et al., 2007). Regarding pass rates, students within the lowest two bands of DELNA were approximately three times more likely to fail compared to their peers. Elder, Bright et al. (2007) found the rate of failure for these students was 23.5%, much higher than that of the undergraduate population as a whole (approximately 10%). Erlam and Botelho de Magalhaes (2021), who analysed correlations between students' DELNA results and their first year GPAs ($N = 5,469$). A significant, modest correlation of 0.33 (a variance of approximately 11%) was found between DELNA's Screening component and GPA. The GPA outcomes of students identified as having below satisfactory academic language skills were skewed towards achieving lower GPAs, suggesting this group of students were at increased risk of failing their subjects in their first semester. The same analysis was conducted for students scoring above the threshold score, with this group achieving higher GPA outcomes. These studies highlight the positive role played by early assessment mechanisms in predicting performance.

Measuring the Academic Skills of University Students (MASUS) has proven to be another well-researched PELA, with empirical studies conducted using the measure since the mid-1990s. Most studies have found a relationship between MASUS and academic performance (Erling & Richardson, 2010), yet there are inconsistencies regarding which MASUS criteria correlate with academic performance. Holder et al. (1999) and later Jones, Holder et al. (2000) found three of four MASUS criteria were significantly correlated with Bachelor of Pharmacy students' ($N = 506$ and $N = 463$) time taken to graduate (i.e., without needing to repeat any subjects). In comparison, Jones, Krass et al. (2000) found that the

overall grades of students in their first year were found to be predicted by just one criterion of MASUS, C: Control of academic writing style. Donohue and Erling (2012), which used MASUS and analysis of feedback on students' assignments from three subjects, as well as student interviews, found that the PELA scores correlated with marks on assignments, ranging from 0.32 and 0.78. One criterion, A: Use of source material, strongly correlated with assignment marks, yet the criteria measuring language (e.g., criterion C: Control of academic writing style and criterion D: Grammatical correctness) did not, which contrasted with Jones, Krass et al. (2000). This was notable, as it raised questions regarding the role of ELP in academic performance. Overall, research outcomes indicate MASUS reliably identified which students required skill development and their strengths and weaknesses regarding language yet were unable to pinpoint the areas in which students may require support.

The "post-entrance literacy assessment", utilised at the University of Notre Dame with students undertaking Health Science degrees (Anderton et al., 2016; McNaught & Shaw, 2016), has also been used to predict academic performance. To date, two predictive studies have been conducted with mixed results overall. First, Anderton et al. (2016) investigated which factors, both academic (e.g., prior academic performance, ATAR scores) and non-academic (e.g., gender, socio-economic status), effectively predicted academic success. Findings indicated that amongst the first-year health science students ($N = 414$) those whose writing was considered sufficient, as measured by the PELA, performed significantly better over the three years in contrast to those in the not sufficient category. In terms of reading, strong performance on the PELA was associated with higher academic performance. Second, Hoyne and McNaught (2016) examined enrolment and performance data of students and their scores on the PELA. Regarding the PELA's ability to predict students likely to struggle, Hoyne and McNaught (2016, p. 7) described it as a "blunt instrument", as "some students who fall below the benchmark score in the PELA ... will go on to pass all units with Credit, Distinction or High Distinction grades." It was hypothesised that this was due to students not being sufficiently motivated when taking the PELA and not completing the task to the best of their abilities.

The final PELA used to predict academic performance is James Cook University Singapore's PELA. Established in 2012, the PELA is compulsory for all commencing first-year students enrolled in their first trimester (Wong et al., 2017). Wong et al. (2017) explored the ability of the PELA to predict the written and academic performance of international

business students in a single business faculty subject. Findings revealed there was a moderate correlation between PELA scores and performance in the subject's assessed paragraph writing task (written performance) and overall grades in the subject (academic performance) with correlation coefficients of 0.14 and 0.33 respectively (Wong et al., 2017), equivalent to variances of between 2% and 11%. Even though the variances reported were lower than other PELA contexts, the researchers maintained that "PELA is a good predictor of students' English writing performance in a business subject" (Wong et al., 2017, p. 9). This study adds to the literature on international PELAs.

In summary, PELAs have been used in predictive studies to determine a relationship between language ability and academic performance. To date, the two PELAs that have received the most attention from researchers has been DELNA and MASUS. When examining findings regarding DELNA and MASUS and academic performance, it appears there are inconsistencies with regard to the precise role ELP plays. For DELNA, results typically predicted approximately 10% of students' performance at university. For MASUS, overall scores displayed inconsistent correlations with academic performance; variances ranged from 10%, with a high of 60% for one assignment. For individual criteria, smaller variances of between 2% to 5% have been determined. Studies using Notre Dame University's PELA indicated that the PELA played an important role in students' academic performance. In addition, research into James Cook University Singapore's PELA revealed variances of between approximately 2% and 11%, not dissimilar to the findings of MASUS studies. These studies highlight the contribution of using PELAs as part of a programme of support assisting in retention and success of students.

2.5 CONCLUSION

Academic performance predictions and the reasons for student attrition at tertiary level are both subjects of developing research. Throughout the literature, academic factors including prior academic performance (e.g., in secondary schooling) and a plethora of non-academic factors (e.g., age, gender, language background) are commonly investigated when predicting university performance. This chapter has reviewed studies on academic factors, including the prior academic performance, usually measured by TER and/or GPA. It has also examined non-academic factors including age, gender, language background, and ELP.

Previous academic performance has proven to be an important factor. Performance at secondary school is considered a sound predictor of academic performance at tertiary level in both Australian and international studies. In comparison, performance at early stages of university has demonstrated a positive relationship with performance later in students' studies. Regarding non-academic factors, it is important to note that although there is agreement among scholars regarding some factors' relationships with academic performance, there is less agreement on certain variables due to universities employing different recruitment and admissions policies, such as types of students targeted and entry score thresholds set (Burton & Dowling, 2005; Grebennikov & Shah, 2012; Harvey & Luckman, 2014).

The inconsistent findings suggest only focusing on one variable may be pointless, due to the plethora of factors at play. For example, entrance scores do not capture all student characteristics (Danilowicz-Gösele et al., 2014) and language background does not provide enough information about relevant student characteristics (Fox, 2004). It is, thus, necessary to identify and monitor patterns of performance, particularly attrition patterns. From the literature reviewed in this chapter, early warning systems of students who may be at-risk of failing and discontinuing their studies are important in terms of mitigating the risk of attrition and enhancing not only retention, but also success rates.

The current thesis has a specific interest in the relationship between students' ELP and their subsequent performance at university. The literature is divided as to the extent language ability can predict performance (e.g., Bridgeman et al., 2016; Elder, 1993; Graham, 1987). This appears to be the case for high stakes language assessments, such as IELTS, administered pre-entry, and PELAs, utilised post-admission. Typically, correlations of approximately 0.30 have been determined, which means ELP can explain roughly 10% of students' academic performance. Due to the lack of clarity regarding ELP and academic performance, universities would benefit from conducting contextualised research to better understand the relationship.

Given the potential for using post-entry measures as predictors, it is important to examine PELAs in greater detail. Chapter 3 reviews the literature on post-admission assessments, starting with a definition and justifications for their use. It then analyses the salient characteristics of PELAs and examines important considerations when administering them. The key consideration of validity is identified with reference to a validity framework.

CHAPTER 3: PELAS AND VALIDITY

3.0 INTRODUCTION

This chapter reviews the literature on post-entry language assessments or PELAs. It begins by presenting an overview of the measures, including a definition, then presents justifications for their use. The chapter subsequently analyses common characteristics of PELAs with reference to four established PELAs: Diagnostic English Language Assessment (DELA), Measuring the Academic Skills of University Students (MASUS), the Diagnostic English Language Needs Assessment (DELNA), and the Post-entry Assessment of Academic Language (PAAL). The chapter then discusses the key considerations of PELA processes. Validity, the most important consideration, is then detailed, including contemporary views of validation studies through the lens of an argument-based approach. The chapter finally examines a framework for PELA validation. At each stage of the framework, reference is made to the framework's application with two PELAs, DELNA and PAAL.

3.1 POST-ENTRY LANGUAGE ASSESSMENTS (PELAs)

Internationalisation and massification, the two major trends in tertiary education over recent decades (Grebennikov & Skaines, 2009; Murray, 2018; Walker-Gibbs et al., 2019; Webb, 2014), have led to questions concerning whether students have the requisite language skills to succeed at English-medium instruction universities. English proficiency tests, such as IELTS, play important roles in determining whether students meet threshold ELP standards prior to commencing their studies. However, as discussed in Chapter 2, obtaining desired scores on a high stakes test does not guarantee success (Pilcher & Richards, 2017; Read, 2019). Murray (2016, p. 121) supports that, "there will likely always be a proportion of students who successfully enrol in degree courses, but who are at risk linguistically." As a consequence, universities have addressed this issue by administering PELAs, post admission.

In terms of a definition, a PELA is any form of language assessment used to identify students already admitted into university who may possess language development needs (Barthel, 2017). These needs are such that they may be at risk of academic failure or not achieving their full academic potential. Read (2015b, p. 217) defined a PELA as a language assessment "administered to students after they have been admitted to a tertiary institution, with a view to identifying those who are likely to struggle to meet the language demands of

their degree programme.” Therefore, the aim of a PELA is to determine students’ academic English skills after admission into university, so that students requiring it can be directed to appropriate support mechanisms. For the purposes of this thesis, a PELA is defined as any form of language assessment administered to students post-admission to identify those who may need subsequent language support. It is important to emphasise that PELAs should not and are not used for the purpose of university admission or exclusion. As noted by Read (2016b), regardless of how poorly students perform on a PELA, they will not be excluded from the degree programme they have been admitted into and enrolled in. This feature makes PELAs low stakes in contrast to the high-stakes nature of pre-admission language measures. Furthermore, a PELA is not an end in itself (Read, 2015a, 2016a). A PELA alone will not develop students’ language proficiency; it must be connected to language enhancement initiatives (Dunworth, 2013b) to overcome the challenges students may face. As argued by Dunworth (2009, p. A-6), “without such tailored programs within institutions the value of a PELA would be diminished.” PELAs are, therefore, an important aspect of a multifaceted approach (Lockwood, 2013) and when connected with appropriate and effective on-campus support, PELAs provide substantial benefits to students.

Another feature referred to when defining a PELA is the process of its application. The PELA process, including follow-up support procedures, was captured by Knoch et al. (2016), who set five minimum parameters for a PELA. A PELA must first target all potential at-risk students to minimise the chance some students may be unsuccessful at university. Secondly, a PELA must capture necessary data with the aim of identifying students’ specific language needs. Thirdly, the information about students obtained via a PELA must be disseminated to the appropriate stakeholders in a manner that is ethical, meaningful and timely. Fourthly, a PELA must be attached to opportunities to address identified language requirements on campus. Finally, consideration of whether the PELA fulfills its pre-defined purpose (e.g., improving the educational outcomes of students) must occur. Examining Knoch et al.’s (2016) minimum requirements for PELAs demonstrates that the assessment itself is only part of the process. Therefore, a PELA must not be considered to be an end in itself; that is, the assessment alone will not assist in developing students’ academic language skills (Read, 2015a, 2016a). No PELA is a “silver bullet” or “panacea” that can overcome contemporary challenges faced by tertiary institutions (Read, 2015a, p. 221). This leads to the imperative that PELAs must be connected to universities’ language enhancement initiatives to overcome the language challenges students face. Such initiatives include consultation with

advisors at support units, such as ALL centres, language developments courses and academic skills workshops.

From a historical perspective, PELAs' prevalence dates back to the early 1990s when the University of Melbourne's Diagnostic English Language Assessment (DELA) was established (Ransom, 2009). Since then, numerous studies have documented or at least mentioned the use of PELAs, including an Australian Government Office for Learning and Teaching funded project by Dunworth et al. (2013a). This project led to the creation of a website, the Degrees of Proficiency, which provided those involved in assessing and developing students' ELP with a variety of resources, including suggestions in PELA implementation. In addition to the Degrees of Proficiency website, several studies have investigated the development, implementation and evaluation of university PELAs (e.g., Read & von Randow, 2013), including DELA (Ransom, 2009). However, the Diagnostic English Language Needs Assessment (DELNA) at the University of Auckland in New Zealand has received the most attention in the PELA literature. Established in 2002, DELNA is considered "the most comprehensive assessment programme of this kind" (Read, 2016b, p. 6). A number of studies have documented DELNA, including the process associated with the PELA and its developments over time (Erlam & Botelho de Magalhaes, 2021; University of Auckland, n.d.b). These ground-breaking PELAs and two additional PELAs illustrate characteristics of the measures in later sections of this chapter.

As far as prevalence is concerned, most Australian universities have embedded some form of post-admission language assessment (Baird & Dooley, 2017; Read, 2019; Veitch & Johnson, 2022). In 2017, a survey was administered to ALL staff employed at 39 Australian universities (Barthel, 2017; Read, 2019). In total, 33 of the 39 Australian universities (85%) responded to the survey, with 70% of the university representatives stating that a PELA was utilised at their institutions. This marked an increase of 8% compared to the findings of a similar survey conducted in 2011 (Barthel, 2017). These statistics, which may be understated (Read, 2019), illustrate the prevalence of PELAs in the Australian university context. Although prevalent, the assessments are very often localised and institution specific (Knoch et al., 2016; Read, 2019; Veitch & Johnson, 2022). Analysing the characteristics of PELAs will, therefore, enable understanding of their use at different universities.

3.2 BENEFITS OF IMPLEMENTING PELAS

This section analyses the benefits of implementing PELAs. These advantages positively impact the following three stakeholder groups: students; university support providers, such as academic language and learning (ALL) centres; and the university as a whole.

Allowing for the identification of students who may be at risk of academic failure at university is the first and foremost benefit of administering PELAs. Most importantly, PELAs when implemented early in students' degree programmes, provide an early measure of students' academic language skills (Fenton-Smith & Humphreys, 2015; Murray, 2016). Early identification of language issues allows plenty of time for development of academic language skills. Read (2015) argued that this ability to identify students potentially at risk of underachievement early in their studies leads to the provision of appropriate feedback and support. Consequently, a direct benefit to students is PELAs' ability to assist in the response to their linguistic needs, which potentially leads to positive impacts on retention and success (Dunworth, 2009; Murray, 2016; Tynan & Johns, 2015). PELAs may also act as a reminder to students of the importance of continued language and self-development, allowing students to take control, whilst potentially leading to increased graduate employability (Baird & Dooley, 2017; Dunworth, 2009).

Another advantage of PELAs is that they introduce available academic language support mechanisms in place at the university (Murray, 2016, 2018). Pointing students to appropriate and targeted support can be critical in terms of student retention and success. Once a student has been identified as requiring additional support, the next stage is promotion and delivery of appropriate and effective language support. The PELA is, therefore, the 'conversation starter' bringing together students and support services, such as academic language and learning units. Importantly, support must be proffered in an equitable manner, as fair access is necessary to ensure effective and efficient use of constraining factors, such as budgets and limited available resources (Murray, 2016, 2018). Such constraining factors continue to be of paramount importance in a post-Covid-19 setting (Barthel, 2023). As argued by Murray (2016, p. 124), PELAs serve as "a filter that helps make certain that resources go to those who stand to benefit most from them, and who might otherwise struggle to cope with their course demands and thus possibly exit their studies prematurely."

A further advantage of PELAs revolves around the beneficial outcomes for institutions themselves. Data obtained via PELAs can assist in identifying areas within the university's educational services that require consideration, including decision-making around curriculum development, resourcing and assessment practices (Dunworth, 2009). Similarly, such data allows institutions to assess their pathways and gatekeeping processes (Dunworth, 2009; Read, 2015a), especially when administered to all targeted students, thus assessing on a "level playing field" (Murray, 2016, p. 123). Additionally, the use of PELAs contributes to the reputation of institutions (Dobinson & Dunworth, 2009; Dunworth, 2009). It assists in the process of quality assurance and compliance with governing body standards, such as TEQSA, demonstrating that the institute is responsive to students' needs and takes educational and ethical responsibilities seriously (Murray, 2016). This reassures already enrolled students about the institution's commitment to student support and may provide a sense of security to potential students. Murray (2016) expressed:

... if a university is responding effectively to the educational and pastoral needs of its students and can demonstrate that it is doing so by identifying their language needs and, where necessary, responding appropriately, then it seems only right that it should be judged favourably on that basis and benefit reputationally accordingly. (p. 123)

Subsequently, PELAs allow investigations into areas such as predictive studies (i.e., to what extent the measure can predict future academic performance) to be carried out using the range of data available (Dobinson & Dunworth, 2009). This ensures that monitoring students after they have commenced their studies can occur and assist in identifying opportunities for support or pathways requiring further investigation (Dobinson & Dunworth, 2009). This can be achieved in a relatively low-cost manner, making it an effective and efficient method of determining the academic English skills of large cohorts of students (Baird & Dooley, 2017; Tynan & Johns, 2015), whilst contributing to increased retention and success rates.

In summary, the advantages of implementing PELAs include benefits for students, university support providers, and the university as a whole. For students, PELAs allow for identification of areas for improvement and leads to assistance regarding their academic and linguistic needs. PELAs also remind students of the importance of continued development and potentially leads to employment outcomes. For support units, PELAs introduce students to the on-campus support options available to them. The promotion and subsequent delivery of academic language support is integral to the effectiveness of the PELA process. Finally,

for universities themselves, PELA data may assist in decision-making around curriculum development, resourcing and assessment practices. PELAs may also enhance the institution’s reputation by underlining its compliance with governing body standards and conveying the message that it is responsive to students’ needs. Overall, university stakeholders benefit from the development of effective PELAs connected to appropriate support mechanisms. Having outlined the benefits of implementing PELAs, the next section analyses characteristics of PELAs to better understand how they are utilised in different contexts.

3.3 FEATURES OF PELAS

Different institutions have adopted various approaches to post-admission language assessment. PELAs in the Australian and New Zealand contexts share common themes in the discourse used to describe them and the principles upon which they are based (Murray, 2016). Common, salient features appear, yet different institutions’ PELAs display distinguishing features. The varied approaches taken presents a challenge in terms of gathering information on the characteristics of PELAs (Read, 2019). However, a survey completed by academic language and learning (ALL) staff employed at 33 of the 39 Australian universities (Barthel, 2017; Read, 2019) sheds light. Moreover, PELA case studies documented within the literature provide further analysis of features. These case studies include: MASUS (University of Sydney), DELA (Melbourne University), DELNA (University of Auckland), and PAAL (University of Melbourne). A detailed analysis of these PELAs’ features can be found in Appendices A to D.

The following sections analyse the PELA features as documented throughout the literature. A matrix of these features and associated questions is shown in Table 3.1 below. The features include PELA targets (i.e., university wide or within schools, faculties), mandatory or optional completion, their developers, the skills assessed, mode of delivery (e.g., pen and paper or online), raters, and links to on-campus support.

Table 3.1

PELA features and key questions

Features	Key Questions
1. Targets	Who is targeted?
2. Mandatory or optional completion	Is the completion of the PELA compulsory for targeted students?

Features	Key Questions
3. Development	Where was the PELA developed and by whom?
4. Assessed skills	What are the skills assessed by the PELA?
5. Mode of delivery	What is the mode of delivery of the PELA?
6. Raters	How is the PELA assessed and by whom?
7. On-campus support	What support mechanisms is the PELA linked to?

3.3.1 TARGETS: ALL OR SELECTED STUDENTS

The first feature of PELAs is their targets. Many universities have elected to administer PELAs to all students from particular cohorts or faculties. Some institutions target pre-identified groups of students from particular cultural or language backgrounds.

The majority of universities that administer a PELA target all students. Overall, 52% of universities employ a PELA institution-wide (Barthel, 2017; Read, 2019). For instance, DELNA, which was developed at the University of Auckland and utilised at various Australian universities, targets all first-year undergraduate students and commencing Higher Degree Research candidates (Read, 2019; Read & von Randow, 2013). Additionally, PAAL was developed at the University of Melbourne to allow post-admission assessment of all students, not just EAL/D students (Elder & Read, 2015). In comparison, MASUS, at the University of Sydney and employed within other Australian university contexts, can be administered to both undergraduate and postgraduate cohorts. However, it is not a university wide assessment, as “faculties can decide whether they would like to implement the procedure with their students” (Knoch & Elder, 2016, p. 219). When decision makers elect to use MASUS with cohorts or faculties of students, however, all students are likely targeted. Targeting all students is notable, as it attempts to ensure each student requiring language support can be identified. This is a positive feature, as English speaking-background students deemed proficient in English may exhibit features that do not adhere to traditional expectations in academia (Murray, 2016; Read, 2016b; Read & von Randow, 2013).

As far as PELAs that target select groups of students are concerned, there is limited insight into precisely which cohorts are commonly targeted. Due to budget constraints, it may be expected that the primary target may be EAL/D students. This group of students has been identified as potentially linguistically ill-equipped to succeed at university and in the workplace (Birrell, 2006; Birrell et al., 2006), and such concerns have resulted in media reports questioning the standards of universities (Baker & Carey, 2019; Dodd, 2019).

Associations between low level language proficiency and academic performance have also been well documented in the literature (Knoch et al., 2016). For EAL/D students, issues in terms of meeting the conventions of academic writing may be apparent, for instance, due to unfamiliarity with academic genres. As an example, DELA at the University of Melbourne targets both undergraduate and postgraduate students (LTRC, n.d., para. 1) who are potentially at risk of academic failure. This group of students includes students entering with minimum required English test results and others entering through non-traditional pathways, such as sub-bachelor foundation courses. DELA is available to all students at the university (Knoch & Elder, 2013); however, it appears the target of DELA is for EAL/D students.

Overall, the analysis of survey data and case studies reveals most PELAs are administered to all students at the university administering the PELA. This means that both English-speaking background students and EAL/D students complete the post-admission assessment. In contrast, some PELAs have been designed to be used with specific groups of students, typically based on their language or cultural background. Administering a language assessment to EAL/D students or students who have been identified as facing linguistic challenges makes sense. However, much of the literature points to the necessity of ensuring all students who may struggle academically are identified, hence the majority of PELAs are used with all students, often throughout the university.

3.3.2 COMPLETION: MANDATORY OR OPTIONAL

The second feature of university PELAs is whether their completion is mandatory or optional for targeted students. Some universities require students to complete a PELA as a compulsory task, whilst others set the assessment as optional, allowing students to decide whether to complete it.

The majority of universities administering a PELA set the assessments as a mandatory task. According to the survey of ALL advisors, slightly more than half (52%) of the PELAs in the Australian university context are compulsory (Barthel, 2017; Read, 2019). Furthermore, analysing PELA case studies illustrates that most are mandatory for targeted students. For DELNA, first-year undergraduate students and HDR candidates are obligated to complete the PELA (Read, 2015a). MASUS completion is also compulsory for cohorts and/or faculties that have been targeted (Knoch & Elder, 2016). DELA, similarly, is mandatory for targeted students, including “international students, and those [domestic

students] entering the university via non-traditional pathways, such as Foundation courses” (Knoch & Elder, 2013, p. 62). Mandating PELA completion on a level playing field is advantageous, as no student is missed, and nothing is assumed (Murray, 2016). However, decision makers must carefully consider achieving compulsory completion. Mandatory PELAs should be presented to ensure positive outcomes, often referred to as positive washback in language assessment literature (Hughes, 1991). For instance, DELNA is introduced to students in a positive manner with emphasis on enhancing educational experiences (Read, 2015a). There will always be a percentage of students who do not complete the task. The mandatory policy surrounding DELA was described by Elder and Read (2015, p. 33) as “patchy” due to communication by various sources within faculties as opposed to being centralised (Ransom 2009). Hence, mandatory PELAs must be designed to eliminate confusion as to their purposes, including the necessity for completion (Read, 2016a). Compulsory PELAs tend to have some form of sanction associated with non-compliance (Read, 2016a), and such sanctions must be also considered carefully.

In contrast to mandatory completion, universities elect to offer PELAs as non-compulsory tasks. Typically, this includes varying levels of persuasion (e.g., “you don’t have to do it, but you will benefit from completing it”). For example, PAAL is not compulsory although students are strongly encouraged to take the test. The task is recommended to all incoming undergraduate and postgraduate students, who may decide where and when they complete the assessment (Knoch et al., 2016). The decision to make a PELA optional is based on academics’ and their institutions’ reluctance to require students to complete an additional assessment of their language skills when they have already successfully matriculated into university (Read, 2016a). The argument for making a PELA optional for targeted students (i.e., all or specific groups of students) is, therefore, that it may be unreasonable to require students to sit a language test after meeting the university’s entry requirements (Murray, 2010; Read, 2016a). Furthermore, allowing students to decide whether or not they complete a PELA places them in a position to take responsibility for self-development (Murray, 2014), as per the English Language Standards for Higher Education (TEQSA, 2012). However, students may have feelings of uneasiness and be unwilling to sit a non-compulsory test. This is evidenced by low PELA completion rates for contexts in which PELAs are optional (Fox, Haggerty et al., 2016; Knoch et al., 2016; Ransom, 2009), which may, in turn, may deprive certain students of ALL support (Murray, 2014).

As can be seen from the above analysis, once targets for completion are determined, approximately half of university PELAs are mandatory. Despite an absence of any real ramifications for non-completion, making a PELA mandatory may increase the chances of identifying students who may require support. Presenting the task on a level playing field also enhances the equitability of the assessment. On the other hand, the resource intensive nature of mandatory PELAs, as well as assessment policy of various institutions, plays a role in decision making in this regard. The debate helps to explain the lack of consistency in terms of applying either a compulsory or optional approach to PELA completion.

3.3.3 DEVELOPMENT: INTERNAL OR EXTERNAL

A third feature differentiating PELAs is the place of their development. PELAs may be developed internally, within the university in which the assessment is utilised. In contrast, universities may enter into an agreement to utilise a PELA created externally.

Designing and developing a PELA internally has become the norm. Notably, the majority of Australian PELAs (83%) were developed within the university where they are used, most often by ALL professionals (Barthel, 2017; Read, 2019). Examining the four PELA case studies, three were created internally; one was created internally with subtests licensed externally. DELA was developed by professional language assessors from the Language Testing Research Centre (LTRC) at the University of Melbourne (Elder & Read, 2015; Knoch & Frost, 2017; Ransom, 2009). In comparison, PAAL was developed by assessment experts at the LTRC to address limitations of DELA, such as its target being particular groups including EAL/D students (Elder & Read, 2015; Knoch et al., 2016). MASUS was also developed in-house, but not by language assessment experts. MASUS was developed by academic language and learning staff (Bonanno, 2002). DELNA shares a similar history; the PELA was created by language assessment experts, including an Associate Professor in Language Studies who had been a language assessment researcher at the University of Melbourne's LTRC, involved in the development of DELA (Read, 2015a; Read, 2016a; Read & von Randow, 2013). However, it differs slightly to other PELA contexts, as subtests are licensed from the University of Melbourne's DELA under a contractual agreement. In-house development, funded by the university, has the advantage of being cost-free for students, as opposed to high-stakes, commercial language tests. Moreover, internal development allows PELAs and the associated processes to respond to specific, contextualised needs of students (Knoch & Elder, 2013).

In contrast to internal development, a small percentage of PELAs were developed externally (Barthel, 2017; Read, 2019). As noted, DELNA's subtests are licensed from the University of Melbourne's DELA under a contractual agreement. Further instances exist for which universities have collaborated, bringing their expertise together to create a PELA (Read, 2016a), including at the University of South Australia, which licensed a PELA developed at the University of Melbourne's LTRC (Knoch et al., 2016).

In summary, due to the ability to fund the development and implementation of PELAs and to respond to the contextualised needs of students, most PELAs have been developed internally within the university at which they are administered. In-house designed PELAs may lack professional validation (Knoch & Elder, 2013). This is a notable concern, considering that universities should present claims of integrity for all assessments regardless of the stakes. Exceptions exist, as is the case at the University of Melbourne and the University of Auckland with DELA and DELNA respectively being developed by language assessment experts and undergoing professional, published validation. However, there is a distinct lack of research documenting evidence for the efficacy and validity of most university PELAs, presenting a substantial gap in the literature, which raises concerns.

3.3.4 SKILLS ASSESSED: MACRO SKILLS, GRAMMAR AND VOCABULARY KNOWLEDGE

An additional feature which differentiates PELAs involves the skills assessed. These skills can be categorised as macro skills (i.e., speaking, listening, reading, writing) and/or skills comprising language knowledge, such as grammatical accuracy and lexical resources or knowledge of vocabulary.

The vast majority of PELAs target students' writing skills. Reading comprehension is commonly assessed, yet listening and, particularly, speaking are neglected. Overall, 83% of Australian PELAs assess writing, 61% assess reading, 13% assess listening, and just one institution's PELA (4%) assesses speaking (Barthel, 2017; Read, 2019). When analysing the four PELA case studies, although writing and reading are commonly assessed, knowledge of vocabulary and listening are also typically included. DELA comprises listening, reading and writing sub-tests (Elder & Read, 2015). In comparison, DELNA is carried out in two stages: a Screening stage, comprised of a 27-item vocabulary test and 73-item cloze-elide (Read, 2015a) assessing vocabulary and reading respectively; and a Diagnostic stage for students requiring it, consisting of listening, reading and writing (University of Auckland, n.d.a).

PAAL shares a number of similarities, containing two screening tasks, a cloze-elide task (similar to DELNA), and a C-Test made up of a series of 100-word texts that have letters at the end of every second word missing (Read, 2015a). It also includes the same writing task as DELA. In contrast, MASUS only assesses discipline specific writing, based on numerical, visual and textual data associated with students' field of study (Bonanno & Jones, 2007; Elder & Read, 2015a), known as an integrated writing task (Plakans, 2012). Consequently, the majority of PELAs target students' writing skills first and foremost, followed by reading, knowledge of vocabulary and to a lesser degree, listening skills. Such skills are typically assessed, as they are the most prevalent for university assessment (Webb et al., 1995). Furthermore, due to the interconnectedness of writing and reading, issues in one skill, such as a mismatch between academics' expectations and students' task interpretations and lack of awareness of how to structure and develop texts (Webb et al., 1995), may affect the other. Regarding writing in particular, sub-standard skills may place students at risk of academic failure (Erling & Richardson, 2010; Lillis & Scott, 2007; Paton, 2007). Hence, it is understandable that most PELAs focus on writing and reading.

In contrast, speaking is seldomly assessed as part of the PELA process. The reason that most universities do not include a speaking assessment, and few include a listening task, is due to practicality. Practicality considerations include the resources needed to develop and administer appropriate speaking and listening tests at scale. Furthermore, research has indicated that speaking skills are less frequently used in comparison to the other three macro skills and is considered less critical from an academic success perspective (Gravatt et al., 1997; Read, 2015a). This point may require consideration as new assessments tasks are developed in response to the rise of Generative Artificial Intelligence, however.

To summarise, examining the survey data and the PELA case studies demonstrates that most PELAs target students' writing skills. Each PELA case study has a writing component, and one PELA, MASUS, comprises an integrated academic writing task. PELAs commonly comprise reading comprehension, as well as vocabulary knowledge, yet listening and, particularly, speaking are neglected. The reasons for targeting writing and reading include the emphasis placed on them within university assessment.

3.3.5 MODE OF DELIVERY: ONLINE OR PAPER-BASED

Another important feature of PELAs is the mode of delivery, either online or pen and paper. Certain university PELAs are administered online, some are paper-based assessments, whilst others might be a combination of online and paper-based assessments.

The majority of PELAs are administered online. The ALL survey indicated that 57% of Australian PELAs are web-based (Barthel, 2017; Read, 2019). It is noteworthy that the percentage of online PELAs has increased from 52% since 2011 (Barthel, 2017; Read, 2019), displaying a move away from traditional approaches towards digital assessments. Inconsistencies are displayed when examining the case studies. PAAL is a completely online assessment, making it accessible from anywhere without invigilation (Knoch et al., 2016). Similarly, MASUS can be used flexibly. The PELA can be delivered online or as a paper-based assessment (Knoch & Elder, 2016). For MASUS, the mode of delivery depends on decisions made for cohorts of students by disciplines or faculties (Knoch & Elder, 2016). The shift to online assessments provides multiple benefits. A PELA delivered online, allowing students to complete the task anywhere, anytime, is less burdensome in terms of financial, human and material resources (Murray, 2014, 2018). Online administration also provides the opportunity for automated marking, increasing efficiency of feedback. It may also be unfeasible to coordinate an opportune time for all test takers to sit the PELA at the same time, or in the same place, without the capacity to provide sufficient resources, such as workstations or computer labs. These features explain the trend towards online PELAs.

Concerns have been raised, however, regarding online delivery, which mostly involve academic or test integrity. When students do not sit the test simultaneously, there may be issues in terms of unfair advantage. Students completing a PELA online may have access to other resources, both online and offline, including generative artificial intelligence, which would impact a PELA's integrity, and the test-takers' actual abilities may be questioned. Consequently, there remains a percentage of PELAs that are paper based. In 2017, 57% of Australian PELAs were paper based (Barthel, 2017; Read, 2019). Notably, pen and paper-based PELAs decreased from 71% to 57% compared to 2011 (Barthel, 2017). As an example, DELA is a pen-and-paper based assessment (Elder & Read, 2015), whilst MASUS can be used flexibly. For instance, the task can be implemented during lectures to large groups in lecture theatres, in smaller tutorial groups or outside of class time, when students can decide when and where to complete the assessment (Knoch & Elder, 2016).

Finally, a hybrid approach has been implemented, with some PELAs employing a paper-based and an online component. For instance, a PELA might involve an online screening sub-test, followed by a pen and paper diagnostic assessment. DELNA was an example of this hybrid approach when it included a computer-based screening task and a paper-based diagnosis (Read & von Randow, 2013). Only students scoring below a threshold level on the screening task were required to complete the paper-based diagnosis.

In sum, a trend towards online administration of PELAs can be observed. Research displays mixed findings amongst the four PELA case studies and the survey data indicated a preference for online mode of assessment. Valid concerns have been raised as a result of this trend, particularly regarding academic integrity. It appears there exists a trade-off between practicality and academic integrity. Researchers have maintained that due to the low-stakes nature of the assessments, there is minimal cause for concern (Read, 2019). Murray (2014, p. 330) argued, “there was little motivation for students to cheat”, indicating that security of the assessment was described as “something of a non-issue”. The trend towards online based language assessment post-admission is predicted to continue, thus academic integrity concerns must be considered, particularly in a world of Generative Artificial Intelligence.

3.3.6 RATERS: HUMANS AS EXPERTS/NON-EXPERTS OR AUTOMATED

An additional characteristic differentiating PELAs is how they are rated and by whom. They may be rated by human raters, who may be language assessment experts, language experts (e.g., ALL staff), discipline or domain experts or other staff at the university. Alternatively, PELAs may be electronically and automatically rated. Unfortunately, a dearth of literature exists specifying who, in terms of experience and/or expertise, rates university PELAs and the level of training received.

In the Australian context, human raters commonly assess students’ PELA completions, particularly the writing component when applicable (Read, 2019). Survey data suggests 61% of PELAs are rated by human raters (Barthel, 2017). Typically, the raters are academic language and learning staff who may be considered academic language experts, but not necessarily language assessment experts (Barthel, 2017). Alternatively, raters may be discipline specific or domain specialists, who may lack language and/or language assessment expertise. For DELA, the listening and reading components are assessed using “detailed marking keys” (Knoch & Elder, 2016, p. 215), presumably by human raters (The University of Melbourne, n.d.). It is assumed raters are language experts, but it is unclear whether they

are also language assessment experts. Similarly, acknowledging that “good quality writing assessment depends on consistent judgements made by well trained raters” (Read, 2015a, p. 54), an experienced team of raters with language expertise rate the Diagnostic writing component of DELNA (Knoch, 2007). Furthermore, trained raters score the writing test of PAAL using a three-category, analytic rating scale (Knoch et al., 2016). However, less is known about who the raters are in terms of professional expertise. Due to the flexibility of MASUS, raters vary depending on the cohort or faculty in which it is administered. Raters’ backgrounds and areas of expertise (i.e., discipline experts or language experts) may vary (Bonanno & Jones, 2007; Knoch & Elder, 2016). Regardless of professional background, raters require adequate training to reliably assess students’ skills based on various criteria, such as grammar, vocabulary, organisation or structure, academic style and task response (Read, 2015a).

In contrast to human raters, the receptive skills of reading and listening sub-tests are often assessed automatically via electronic means. According to the survey data, 30% of PELAs are rated automatically (Barthel, 2017). Notably, results from a survey conducted in 2011 indicated a higher percentage (78%) of PELAs were rated by human raters (e.g., ALL staff; Barthel, 2011), yet this figure has decreased to 61%, demonstrating a trend towards automated rating of PELAs. For instance, the DELNA Screening test is automatically scored via computer-aided rating (Read, 2015a). The trend to online, automated rating is likely to continue given the advancements in technology such as Generative artificial Intelligence.

In sum, how the assessments are rated are distinguishing features of PELAs. PELAs may be electronically and automatically rated, yet most are rated by human raters. Human raters may be language experts (e.g., ALL staff), language assessment experts, discipline experts or other staff at the university. Currently, human raters typically assess students’ PELA performance. Raters, the training provided, and moderation practices are critical features of PELAs. However, limited research has precisely pinpointed these features; therefore, further investigation is required.

3.3.7 LINKS TO ON-CAMPUS SUPPORT: CONSULTATIONS, WORKSHOPS OR SUBJECTS

A final, yet critical feature of PELAs is their link to on-campus support. As demonstrated, a PELA is meaningless without linkages to some form of support mechanism provided to students by the university. Support mechanisms, therefore, aim to assist students

in the development of their academic English proficiency, academic literacies, and/or communications skills. There are four main categories of support: one-to-one consultations with advisors or tutors, academic language workshops, credit-bearing subjects, and other mechanisms including online resources and peer assisted mentoring (Murray, 2018; Read, 2016a). Unfortunately, limited publications exist documenting the delivery of these support mechanisms attached to PELAs and investigations into their efficacy are even scarcer.

The first category of on-campus support is one-to-one consultations with advisors, such as academic language and learning staff and English tutors. For instance, support for students completing DELNA is based on the Diagnostic stage results; students scoring below a threshold level are requested to attend a face-to-face consultation with an Academic English Advisor. Writing feedback includes a detailed breakdown in the form of strengths and weaknesses of each criterion: fluency, content, and grammar and vocabulary, with raters' notes making up the feedback given during face-to-face consultations (Read, 2015). Advisors offer suggestions concerning language enhancement opportunities at the University. For PAAL, students' writing is accessible online for English tutors to use for diagnostic purposes (Knoch et al., 2016). It is during such one-to-one consultations that the diagnostic features of PELAs are realised (Read, 2015a; Read, 2019).

The second category of on-campus support is academic language development workshops. Different PELA contexts may offer these workshops, dependent on the faculty administering them. For DELA, subsequent language development options include academic English workshops delivered by the university's ALL unit (Knoch & Elder, 2016). Students in the "at risk" group are required to participate in such language support mechanisms; "borderline" students are suggested to participate in such offering, yet it is not compulsory (Knoch & Elder, 2016, p. 216). Regarding MASUS, the department administering the test makes decisions concerning follow-up actions, which may include "... the introduction of a series of workshops or other support for students deemed at risk in their subject courses" (Knoch & Elder, 2016, p. 220).

The next category of support is credit-bearing subjects, such as English for academic purposes or English for discipline specific purposes subjects. Students would enrol in these subjects and their completion would appear on students' academic transcripts. Of the PELA case studies, DELA is the only PELA connected to for-credit subjects. Examples of such subjects include Academic English 1 for undergraduates and Presenting Academic Discourse

for postgraduates (Knoch & Elder, 2016). Furthermore, Academic English tutorials run over 10 weeks for postgraduate students undertaking science, arts, food science, education, engineering and information technology degrees (Knoch & Elder, 2016).

A final category of support for the purposes of this review is labelled other mechanisms. These include on-campus support such as online resources and peer assisted mentoring, commonly referred to as peer assisted study sessions or PASS (Zaccagnini & Verenikina, 2013). For instance, DELA is connected to a peer-assisted support programme, “Melbourne Talks” (Knoch & Elder, 2016).

In summary, on-campus support linked to PELA results are integral to the efficacy of the PELA process. Without on-campus support, a PELA does not seem to serve any discernible purpose, particularly when it comes to students’ language development. Unfortunately, there is limited published information about the follow-up support mechanisms for PELAs. From the limited literature available, the programmes of support proffered at different universities appears contextualised. Support initiatives may be determined by the department which administers the PELA, individual faculties and/or the institutes via university-wide policy. Concerns have been raised as to the uptake of support, described by Elder and Read (2015, p. 33) as “patchy” due to “very few” of the students requiring support taking up the opportunity for language development (Elder & Read, 2015, p. 34). If support recommendations are not embedded into the content of the subjects, student uptake may not be mandatory nor monitored (Knoch & Elder, 2016). A second concern is that the diagnostic feature of PELAs may not be fully realised. Knoch and Elder (2016, p. 216) argued there is “limited opportunity for students to discuss their results and recommendations or any other follow-up options.” This indicates a potential missed opportunity to receive feedback on their performance on the PELA itself. There is an opportunity to further investigate the role and efficacy of on-campus support mechanisms.

3.3.8 SUMMARY OF PELA FEATURES

The literature reviewed in the previous sections identified the features of PELAs. Overall, seven key features were discussed. The first feature regards their targets. The majority of universities administering PELAs have targeted all students from particular cohorts or faculties. In contrast, some have targeted pre-identified groups of students from particular cultural or language backgrounds. Typically, the primary target may be EAL/D

students. However, PELA literature emphasises the need to identify those who may struggle academically. The second feature of university PELAs is compulsion to complete the assessment. Most universities set the assessments as a mandatory task, although they lack repercussions for non-completion. Some offer the assessment as an optional task, yet these cases often report low completion rates. A third feature is whether they have been developed internally, within the university in which the assessment is utilised, or externally. Creating a PELA internally has become common practice, most often by academic language and learning professionals. Some PELAs have been developed externally with universities licensing the assessment. A fourth feature concerns the skills assessed. Writing and reading skills were identified as the most commonly assessed skills. Grammar and vocabulary knowledge and listening skills were common features of the PELA case studies. Speaking is seldomly assessed. A fifth feature is the mode of delivery, either online, pen and paper or a hybrid approach. The majority of PELAs are administered online; some are paper-based, whilst others are a combination of online and paper-based assessments. Concerns have been raised regarding online delivery, which mostly involve academic or test integrity. A sixth feature of PELAs is how they are rated and by whom. Human raters most often assess PELAs, particularly the writing component if applicable (Read, 2019). Reading and listening sub-tests are often assessed automatically via electronic means, and this is a trend that is likely to continue given the advancements in Generative artificial Intelligence. The final, and arguably most important feature is their link to on-campus support. Four main categories of support were identified, including one-to-one consultations, academic language workshops, credit-bearing subjects, and other mechanisms including online resources and peer assisted mentoring. Without on-campus support, a PELA serves no discernible purpose, Unfortunately, there is limited published information about the follow-up support mechanisms and their effectiveness.

The literature provides some insight into PELA features. In particular, the survey conducted amongst university academic language and learning advisors (Bathel, 2017; Read, 2019). Furthermore, four PELA case studies explain features in the contexts they are administered. Table 3.2 summarises the features of four PELA case studies discussed in this chapter. The next section of the chapter will examine the key considerations regarding PELAs.

Table 3.2*Features of PELA case studies*

PELA	Targets?	Mandatory?	Developers?	Skills?	Mode?	Raters?	Support?
Diagnostic English Language Assessment (DELA)	Students potentially at risk of academic failure	Mandatory	Professional language assessors	Reading, writing, listening	Paper based	Trained raters	Academic English subjects, tutorials, peer assisted support, ALL workshops
Measuring the Academic Skills of University Students (MASUS)	Specific cohorts and/or faculties	Mandatory	ALL staff	Writing	Paper based or online	Language assessment experts or discipline experts	Dependent of faculty
Diagnostic English Language Needs Assessment (DELNA)	All undergraduates and HDR candidates	Mandatory	Professional language assessors	Reading, listening, writing (if required)	Online, paper based	Auto-scored, trained raters	ALL consultations
Post-entry Assessment of Academic Language (PAAL)	All undergraduate and postgraduate students	Optional	Professional language assessors	Vocabulary, reading, writing	Online	Auto-scored, trained raters	English tutors

As illustrated in Table 3.2, although similarities exist, each PELA is unique. The Diagnostic English Language Assessment (DELA) targets students potentially at risk of academic failure, is mandatory, and assesses reading, writing, and listening skills. The Measuring the Academic Skills of University Students (MASUS) is also mandatory and targets specific cohorts and/or faculties, focusing on writing skills. The Diagnostic English Language Needs Assessment (DELNA) is mandatory for all undergraduates and HDR candidates, assessing reading, listening, and writing skills if required. Lastly, the Post-entry Assessment of Academic Language (PAAL) is optional for all undergraduate and postgraduate students and assesses vocabulary, reading, and writing skills. Each PELA, except MASUS, was developed by professional language assessors and rated by trained raters including language assessment experts. The support provided varies from academic English subjects, tutorials, peer-assisted support, and workshops to consultations at academic language learning centres and English tutors.

3.4 KEY CONSIDERATIONS OF PELAS

Having presented the benefits of university PELAs for students, support services and universities, as well as their features, this section examines the key considerations. These considerations include practicality, potential stigma amongst students who perform poorly, the resource intensive nature of PELAs, and the major concern identified in the literature, validity evidence or lack thereof.

The first consideration of PELAs concerns language assessment practicality, which refers to the relationship between the required resources for developing and administering the assessment and the available resources (Cushing Weigle, 2002). The associated work required to develop, administer, rate and communicate PELA results to students can prove considerable (Dunworth et al., 2013a; Knoch & Elder 2013, Murray, 2010, 2016). A number of logistical issues are present in terms of test administration and subsequent procedures (Murray, 2010), with much of the work being tasked by ALL units, who are typically understaffed. Due to this resource intensive nature of PELAs, a critical consideration is achieving the desired purposes whilst adhering to practicality constraints.

A second important consideration regards designing and promoting PELAs in a manner that leads to student uptake of support. The literature on student retention emphasises the necessity of ensuring students' familiarity with support options available (Lobo, 2012),

yet students can be nervous, unaware and/or unwilling to seek support (McLeod, 2012). Unfortunately, very often students requiring support elect not to engage (Murray, 2012; Podorova, 2016; von Randow, 2013). Limited literature documents student compliance with PELA programmes, but the available literature indicates compliance rates vary considerably, particularly when participation is not compulsory (Ransom, 2009). As an example, Fox, Haggerty et al. (2016) reported that between 2011 and 2012, only 2% of undergraduate engineering students who completed a PELA sought feedback on their results. Poor uptake of support could be due to students mistakenly believing that receiving support may jeopardise their academic progress (Elder & Erlam, 2001). It is, hence, imperative that PELAs and associated support are designed and presented to students in a way that conveys their purposes and benefits of engaging.

A potential stigma associated with poor performance, which may demotivate students academically, poses an additional consideration. Dunworth (2009) indicated that PELAs, which may be considered remedial instruments, carry the stigma of failure. If students perform to a below satisfactory standard, they may feel as if they have failed. Murray (2010) concurred, stating students whose communication skills are deemed as inferior compared to their peers may feel marginalised. Given PELAs may be the first assessment students complete, if stigma is associated with poor performance, it may have serious negative impacts. No adverse outcomes of PELA results, therefore, becomes imperative.

A final consideration, which encompasses all of the aforementioned concerns raised regarding practicality, uptake of support and stigma, concerns validity. Given that the majority of PELAs are developed in-house within institutions' ALL units by staff who may not be language assessment experts (Elder & Read, 2015; Read, 2015a), the assessments often lack professional validation (Knoch & Elder, 2013). Another reason to explain this lack of validity evidence is that the validation process is onerous in terms of time and financial resources (Dobinson & Dunworth, 2009). Simply, investigating the validity of a PELA requires much effort. As argued by Knoch et al. (2016, p. 24), due to "the diverse range of students who may experience difficulties with academic English and the variable nature of the difficulties they face, the validity challenges for PELA design and use are considerable". Consequently, a paucity of evidence for their validity has been determined (Knoch et al., 2016), which is concerning (Dobinson & Dunworth, 2009; Dunworth, 2009). Even though validity of a testing instrument must be evaluated as a precursor to addressing

research questions involving a PELA (Read, 2010), how such evaluation can be achieved poses a challenge (Read, 2016b). Encouragingly, a framework exists based on a hybrid of respected argument-based models, which can assist in the presentation of a validity argument for PELAs (Knoch & Elder, 2013). The next sections review the literature on language assessment validity and present the PELA validity framework.

3.5 LANGUAGE ASSESSMENT VALIDITY

This section examines the most significant concern regarding PELAs, their validity. In language assessment, validity is regarded as of critical importance due to decision makers relying on professionals to apply the procedures for investigating and establishing validity of assessments (Chapelle, 2011, 2012; Fulcher, 2015; Fulcher & Davidson, 2012). The section firstly defines validity and validation. It then presents a brief historical overview of validity in language assessment, through the lens of an argument-based approach and the rationale for undertaking validation studies. Next, the section analyses a framework for PELA validation, with reference to how the framework was applied to two PELAs introduced in previous sections, DELNA and PAAL.

3.5.1 BACKGROUND ON VALIDITY AND VALIDATION

Within the domain of philosophical logic, an argument is valid when its conclusions follow on from its premises. In other words, if the argument's premises are true, its conclusion is also true; its conclusion, thus, cannot be false. Hence, consistency between the premises and the conclusion is required (Davies & Elder, 2005). Validity is defined by The Standards for Educational and Psychological Testing, as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p. 11). This focus on interpretations and uses of assessment scores was reflected in Messick's seminal definition of validity. Messick (1989) first opined that validity is "an overall evaluative judgement of the degree to which empirical evidence is and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores" (p. 13). As a consequence, validity entails the demonstration of the meaning of a language assessment score and the justification of the score's use within the context in which it is employed (Chapelle, 2012). Notably, the assessment itself cannot be considered "valid" or "invalid". Rather, it is the interpretations and uses of assessment scores that are the focus of validity studies. Overall,

while researchers agree on the definition of validity, there is less agreement in terms of how to investigate validity (i.e., validation).

Validation can be defined as the process undertaken to obtain evidence to support the claims made about the ability or abilities of an individual (i.e., a university student in the case of PELAs), based on their scores on the assessment (Davies & Elder, 2005). As noted by Cronbach in 1990, the process is “an inquiry into the soundness of the interpretations proposed for scores from a test” (p. 145). Thus, validation studies seek to investigate the extent to which an assessment proffers true representations of individuals regarding the context of its use. These investigations involve the use of logic, judgement, theoretical and empirical evidence to determine validity. Consequently, validity is analysed through the process of validation (Davies & Elder, 2005). In order to achieve clarity regarding validity and, therefore, validation, the evolution of the understandings of these key concepts is examined next.

Validity is not a new concept; theoretical conceptions of validity or “validities” were proposed throughout the last century (e.g., Cronbach & Meehl, 1955; Lado, 1961; Thorndike, 1918). Validity refers to the requirement that an assessment assesses what it claims to assess (Chapelle, 2011). Both the focus of validity being solely on the test itself and the scope of validation studies has evolved since Messick (1989) presented validity as a unitary construct, proposing an integrated view of validity. For the first time, varied conceptions of validity theory were brought together into a single, all-encompassing theory (O’Sullivan & Weir, 2011). Specifically, Messick viewed assessment as a reasoning and evidence gathering process carried out to make inferences about test takers (Knoch, 2007). Subsequently, a shift occurred in understanding validation procedures and their scope. Modern conceptions of validity do not involve discussions about “multiple validities”, such as criterion validity and construct validity, nor do they include placing excessive reliance on one particular aspect of validity. In contrast, the discussion in contemporary language assessment concerns multiple sources of evidence for assessment validity (Davies & Elder, 2005; Messick, 1989). For instance, researchers no longer view reliability as separate from validity, but as an important aspect (Davies & Elder, 2005). Consequently, there has been a shift away from enquiries that encompass only statistical analysis, such as correlational studies of assessment scores, to a focus on the nature of the assessment, the quality of the performances that are elicited during the assessment and the plausibility of the interpretation of test scores and their uses in making

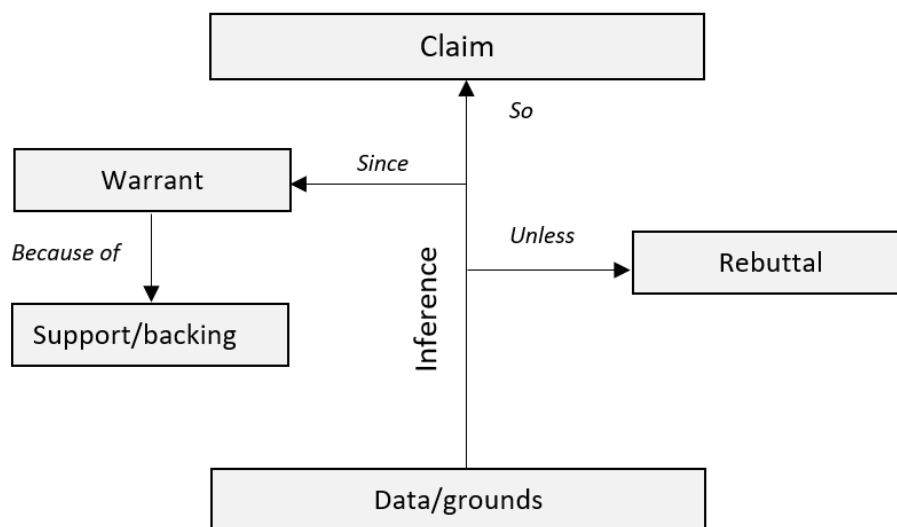
decisions (Davies & Elder, 2005; McNamara & Roever, 2006). This change in philosophy has led to many researchers introducing interpretative arguments to clearly define test score uses and interpretations (Bachman, 1990; Bachman & Palmer, 1996, 2010; Chapelle, 1999). The validation process, therefore, requires gathering a broad range of data that adds to the evaluation of an assessment's validity (Davies & Elder, 2005), known as the presentation of a validity argument.

3.6 ARGUMENT-BASED APPROACHES TO VALIDATION

To provide a clear understanding for how to conduct validation studies, Kane (1992, 2001, 2012) drew on the work of Cronbach (1988) and, in particular, Toulmin (1958, 2003) to present an argument-based approach to validation. The argument-based approach refers to an analytical method used to conceptualise, conduct and interpret validation studies for educational and psychological assessment contexts, including language assessment (Chapelle & Voss, 2021a). To better understand the argument-based approach, it is beneficial to analyse its theoretical underpinnings. Toulmin (1958, 2003) put forward an argument-based model for practical arguments, which involved making claims based on data or grounds from the context of the argument. Such arguments are presented to form a case for a conclusion via a chain of arguments (Chapelle et al., 2008). At its basic level, an argument consists of data, a conclusion that is drawn based on such data and the inference that connects the data and the conclusion (Chapelle & Lee, 2021). In order to advance from one step to the next, conditions are to be met. These conditions are labelled warrants (Toulmin, 1958), and warrants support the inference. Evidence or backing is required to support each warrant. It is necessary to provide support for warrants in the form of theoretical and/or empirical data. If evidence cannot be provided or if the warrant can be refuted, “the chain is broken” and a validity argument is unable to be presented (Roever et al., 2014, p. 61). Figure 3.1 illustrates the argument structure as described by Toulmin (2003).

Figure 3.1

The structure of argument presented by Toulmin (2003)



Note. Adapted from *Building a validity argument for the Test of English as a Foreign Language* by Chapelle et al. (2008, p. 7). Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

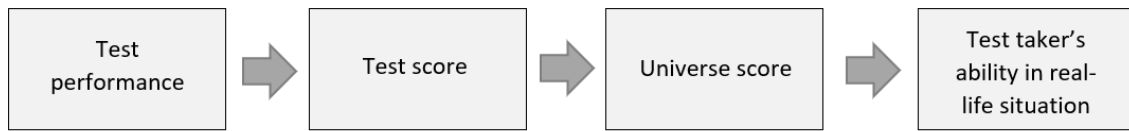
An argument-based approach to validity requires the following components: the claims that are made need to be clearly stated and the veracity of such claims must be evaluated (Kane, 2012). Claims are conclusions that the designer of an assessment wants to draw (Read, 2015a) and are presented regarding an assessment score's meaning and how it will be used (Knoch & Macqueen, 2019). An example claim for a PELA is, "the score on the test is an adequate reflection of the observed behaviour" (Knoch & Elder, 2013, p. 55). Connecting the claims are inferences, defined as propositions test users make based on the observed performances leading to decisions (Chapelle, 2011; Fulcher, 2010). Inferences ought to be outlined creating bridges that connect observed performances to conclusions and, subsequently, decisions based on assessment results (Roever, 2011). The inferences rely on warrants and assumptions, which require evaluation. Warrants are defined as accepted rules, procedures or guidelines (Chapelle et al., 2008; Knoch & Macqueen, 2019) and are based on assumptions. An example warrant for the aforementioned claim for a PELA is, "scoring criteria and rubrics capture relevant aspects of performance" (Knoch & Elder, 2013, p. 56). Each claim contains multiple warrants depending on the context of the assessment. It is when these warrants and assumptions, which underlie the inferences, are found to be credible and plausible that the inferences are supported. Sources of evidence to demonstrate such credibility and plausibility regarding a PELA might include, for instance, statistical analysis of test scores, student questionnaires and review by language test experts. Consequently, the

process of presenting a validity argument for a PELA requires the gathering of a broad range of evaluation data by proceeding through the inferences, claims and warrants at each level of the validity argument (Davies & Elder, 2005). The validity argument that is presented for one PELA may be substantially different to the argument presented for another (Knoch & Macqueen, 2019).

Kane (2012) labelled the presentation of the inferences, and connected warrants and assumptions, the interpretive argument. The interpretive argument, made up of the claims laid out, are evaluated, not in terms of being valid or invalid, but in terms of their plausibility measured by degrees. In other words, “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA, APA, & NCME, 2014, p. 11). The argument must be clearly and coherently stated for claims and assumptions to be conveyed so that conclusions are reasonable. The interpretive argument (i.e., the first step in the process), functions as a framework for the second step, the validity argument. This is because in the second part of the process (i.e., the validity argument), evidence is collated to support inferences and assumptions laid out in the first step (Kane, 2006). The interpretive argument, consequently, assists in identifying what research is required in order to obtain evidence for the validity argument. Figure 3.2 highlights the building blocks of Kane’s (1992) interpretive argument. The interpretive argument is based on the following elements: test performance, test score, universe score, and the test taker’s ability in real life (Kane, 1992; Kane et al., 1999). The test-taker’s response or responses to assessment tasks is the observed test score (Roever, 2011). The observed test score is then generalised as the universe score, in other words “the whole universe of possible items and responses” (Roever, 2011, p. 464). The universe score is subsequently extrapolated as the target score across the target domain or “the range of observations associated with the attribute of interest” (Kane, 2006, p. 31) and is the test taker’s ability in a real-life situation.

Figure 3.2

Building blocks of Kane's 1992 interpretive argument



Note. Adapted from Knoch and Elder (2013, p. 50). <https://doi.org/10.58379/YZLQ8816>. Permission granted from authors/copyright owners.

Between each element of the framework (i.e., test performance, test score, universe score, and the test taker's ability in real life) is an inference which works as a bridge connecting building blocks or individual arguments (Knoch & Elder, 2013). The arrows in Figure 3.2 represent the inferences (i.e., Evaluation, Generalisability, Explanation and Extrapolation), considered bridges between the individual argument (Knoch & Elder, 2013). Each bridge must be supported by evidence. This allows for a systematic and coherent argument to be presented and understood. The bridge, which connects "test performance" and "test score" is the Evaluation inference. The Evaluation inference refers to the "quality of test scores as meaningful summaries of how the students performed in the test" (Read, 2015a, p. 201). Here, support is required to prove, for example, that the assessment is conducted with consistency, that scoring procedures are true reflections of the construct of the assessment, and that the instructions and the administration procedures do not interfere with what the assessment was created to measure. The next bridge, which connects "test score" and "universe score" is the Generalisability inference. This inference assumes that the score that is obtained on the assessment represents the same score that would be obtained on another version of the assessment, irrespective of who was rating the test taker's response. In other words, it is assumed that the assessment is a reliable measure (Knoch & Elder, 2013). Finally, the bridge connecting "universe score" and "test taker's ability in real-life situation" is the Explanation and Extrapolation inference. Explanation connects a test-taker's performance on the assessment with the underlying construct in a way that is more general than the first inference, Evaluation (Read, 2015a). Here, supporting evidence includes whether the assessment elicits processes and strategies as per the intention of test designers. Extrapolation refers to the relationship between the assessment tasks and what is required, in terms of knowledge and skills, to carry out such tasks in a real-world context (Read, 2015a). Evidence includes how accurately the assessment tasks are consistent with the linguistic demands of the appropriate real-world situation or domain.

3.6.1 CRITIQUE OF THE ARGUMENT-BASED APPROACH

There are several advantages of adopting an argument-based approach to validation. The first salient advantage is that it provides guidance in terms of how to organise and synthesise evidence for validation research (Davies & Elder, 2005). It provides guidance for the organisation of a validation study by assisting in synthesising evidence from various stakeholder perspectives and considering counterevidence refuting claims presented about the test (Chapelle et al., 2010). Secondly, it moves away from a “checklist” or “ticking boxes” approach to validity, which was the norm prior to the 1990s. The argument-based approach deviates from overly simplistic validity interpretations (Fulcher & Davidson, 2012; Haertal, 1999). Thirdly, due to the consideration of claims and their associated rebuttals, there is a reduced chance that investigations will focus solely on confirmatory evidence. Rather than merely collecting evidence that confirms the validity of a measure, evidence is sought that highlight problems with the interpretation (Chapelle et al., 2010; Fulcher & Davidson, 2012; Haertal, 1999). This aligns with Cronbach (1980, p. 103), who stated, “the job of validation is not to support an interpretation, but to find out what might be wrong with it.” Finally, the argument-based approach to validation allows the researcher to determine what additional evidence is required to strengthen the validity argument. Any gaps appearing indicate opportunities for future investigation to support specific inferences in the validity argument, leading to recommendations to respond to such weaknesses (Chapelle & Voss, 2021b).

However, the framework and its subsequent iterations were not without limitations. The original model failed to take into account the consequences of test scores (Knoch & Elder, (2013). Kane addressed this limitation in later developments of the model (e.g., Kane, 2001), in which Decisions were included as an inference. However, missing from Kane’s interpretive model were test use and test consequences. In response, Bachman (2005) and Bachman and Palmer (2010) put forth a process of presenting validity arguments based on Kane’s model. The model, labelled “assessment use argument”, included test consequences and placed emphasis on consequences post score interpretation, for instance whether test-takers are required to participate in language development activities, particularly relevant for PELAs, as “the success of any PELA initiative relies on uptake of the advice stemming from test results” (Knoch & Elder, 2013, pp. 52-53). The assessment use argument is advantageous in the context of investigating a validity argument for a PELA, as it incorporates assessment score usefulness and consequences of using scores to make

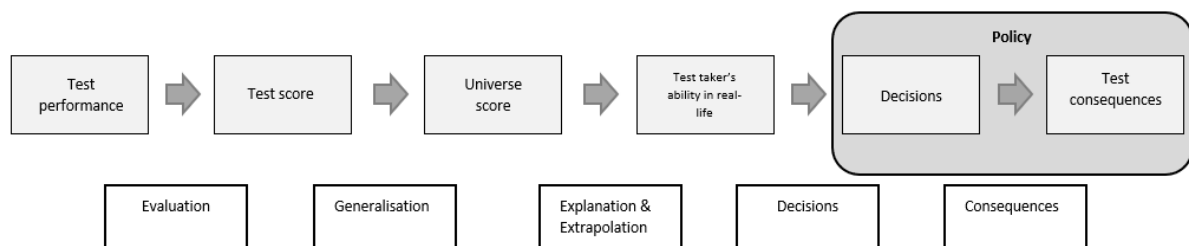
decisions. The following section examines an existing PELA validity framework that incorporates both the argument-based approach and the assessment use argument.

3.7 PELA VALIDATION FRAMEWORK

Achieving a standard of quality for PELAs ensures they fulfill their roles within universities' language policies (Read, 2015a) and assist in student success. PELA users (e.g., students, advisors, academic staff, administrators) must make informed decisions. Knoch and Elder (2013), consequently, introduced a systematic framework detailing the arguments and the requisite supporting evidence underpinning validity claims in the context of PELAs. Based on the works of Kane (1992, 2001), Kane et al. (1999), Bachman (2005), and Bachman and Palmer (2010), the framework contains a series of warrants for each individual inference, applicable to a variety of PELA contexts (Knoch et al., 2016). It details the inferences, claims, warrants and suggested supporting evidence required to present a validity argument for a PELA. The building blocks for this hybrid model are demonstrated in Figure 3.3.

Figure 3.3

Building blocks of the hybrid framework for evaluating PELAs



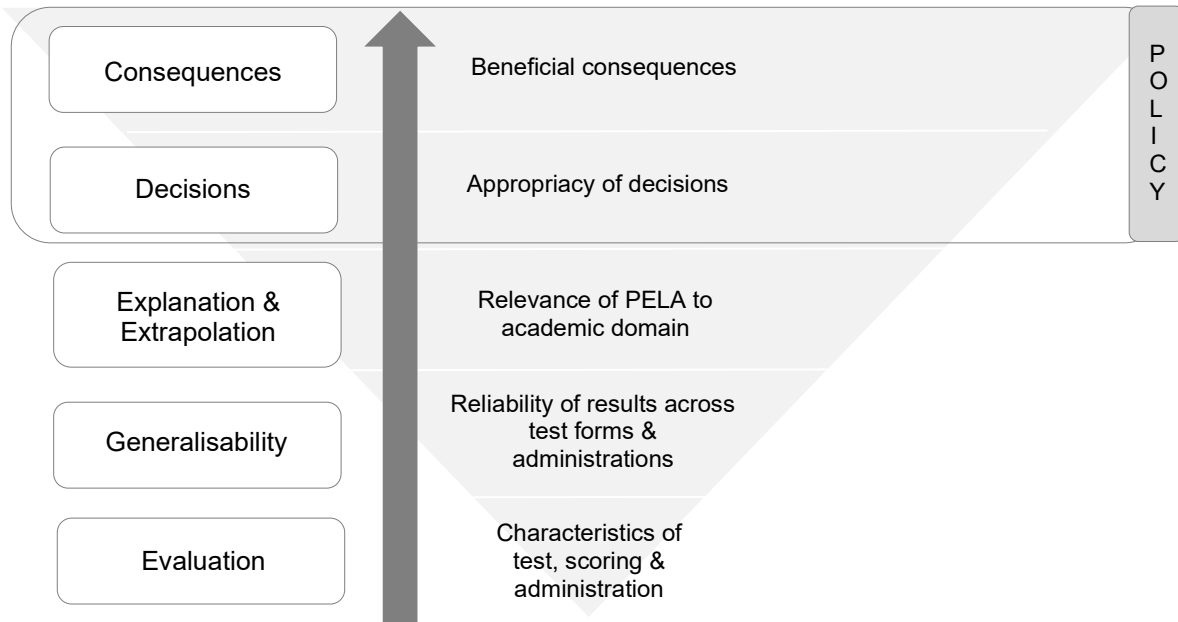
Note. Adapted from Knoch and Elder (2013, p. 52). <https://doi.org/10.58379/YZLQ8816>. Permission granted from authors/copyright owners.

As the figure above illustrates, the PELA validation framework comprises five inferences: Evaluation, Generalisability, Explanation and Extrapolation, Decisions and Consequences. These five inferences are based on the following claims: 1. the PELA score adequately reflects observed behaviour (Evaluation inference); 2. the PELA yields consistent results across various assessment contexts (Generalisability inference); 3. the PELA reflects the language skill construct that is targeted, providing information on the abilities of test takers (Explanation and Extrapolation inference); 4. Decisions based on PELA scores are appropriate and clearly communicated (Decisions inference); and 5. The consequences and decisions made through the use of the PELA benefit all stakeholders (Consequences

inference). Figure 3.4 displays a visual summary of the framework including the associated inferences, as represented by an upside-down pyramid.

Figure 3.4

Overview of building blocks and inferences contained in the framework for PELA validation



Note. Adapted from Knoch and Elder (2013, p. 54) <https://doi.org/10.58379/YZLQ8816>. Permission granted from authors/copyright owners.

As can be seen in the figure, the inference that is the most test focused, Evaluation, appears at the bottom of the pyramid. The Evaluation inference focuses specifically on the reliability of the assessment, assessment items' functioning, how the assessment is scored, and how it is administered. Progressing upwards, the focus becomes the Generalisability of the PELA. That is, the generalisability of the results across administrations, tasks and raters. The focus then shifts to the relevance of PELA tasks within academia, as well as models of academic language ability (Explanation and Extrapolation). Attention then turns from the assessment itself to the area of policy, including Decisions made and Consequences of such decisions. The appropriateness of outcomes based on PELA results comprises the Decisions inference. Finally, Consequences focuses on the outcomes of such decisions and their benefits for stakeholders, including students as test takers.

To demonstrate the framework's application, the following sections analyse the claims made, the warrants for each inference and potential supporting evidence. Each section

focuses on a separate inference and includes a subsection referring to the validity arguments presented for two PELAs, DELNA and PAAL.

3.7.1 EVALUATION INFERENCE

The first inference, Evaluation, focuses on how adequately the assessment score reflects the behaviours of test takers. The claim made at this level is that “the score on the test is an adequate reflection of the observed test behaviour” (Knoch & Elder, 2013, p. 56). The inference is based on the following assumptions as per Chapelle (2008): 1. appropriate rubrics to score responses are used which provide evidence of targeted abilities, 2. test administration conditions are appropriate which provide evidence of targeted abilities, and 3. statistical characteristics of measures, items and different test forms are appropriate to make norm referenced decisions.

The Evaluation inference comprises five warrants. Properties of the rating scale (warrant 1) and the raters’ application of the scale (warrant 2) are the focus of the first two warrants. The third warrant focuses on test administration conditions being clearly communicated and appropriate for the task and the test takers. Similarly, the fourth warrant requires clarity of instruction for test takers. Warrant 5 requires the assessment to be pitched at a difficulty level considered appropriate for the target audience. It also requires tasks and/or items to consistently discriminate to identify more and less able test takers. Table 3.3 below illustrates the warrants and possible supporting evidence to be obtained for the Evaluation inference.

Table 3.3

Warrants and potential supporting evidence for the Evaluation inference

Evaluation inference	
Claim: The score on the test is an adequate reflection of the observed test behaviour.	
Warrants	Sources of supporting evidence
1. Scoring criteria and rubrics capture relevant aspects of performance.	Review by language testing and domain experts.
2. Raters can implement scoring procedures consistently.	Statistical analysis of test scores.
3. Test administration conditions are clearly articulated and appropriate.	Student questionnaires/interviews; review of test administration protocol; observation

Evaluation inference

Claim: The score on the test is an adequate reflection of the observed test behaviour.

	of test sessions, interviews with test invigilators.
4. Instructions and tasks are clear to all test takers	Student questionnaires/interviews.
5. Test is pitched at appropriate difficulty level and test tasks/items discriminate consistently between more and less able candidates.	Statistical analysis of test properties (i.e., item difficulty, discrimination, internal consistency).

Note: Adapted from Knoch and Elder (2013, p. 56). <https://doi.org/10.58379/YZLQ8816>. Permission granted from authors/copyright owners.

As shown in the table, the Evaluation inference focuses on test design aspects, such as statistical data to ensure the test is operating as it should (Knoch & Elder, 2013). Backing for the properties of the rating scale can be obtained through factor analysis, multi-trait multi method analysis or many-facet Rasch analysis, for instance (Knoch & Chapelle, 2017). Rater reliability can be evaluated via Classical Test Theory, many-facet Rasch analysis and Generalizability theory (G-theory; Knoch & Chapelle, 2017). Furthermore, questionnaires employing open-ended type questions can determine how confident raters are in applying the rating scale (Knoch & Chapelle, 2017). Support for the Evaluation inference is demonstrated through evaluating whether scores observed display intended characteristics (Knoch & Chapelle, 2017). It is at this level of the framework where evidence is “often lacking for Australian PELAs or at least not publicly available” (Knoch & Elder, 2013, p. 56). The following section presents how the Evaluation inference of the framework was applied to DELNA (Read, 2015a) and PAAL (Knoch et al., 2016).

This section now presents evidence that formed the backing for each warrant of the Evaluation inference for DELNA, as per Read’s (2015) application of the framework and PAAL, according to Knoch et al. (2016). Warrants 1 and 2 primarily concern the evaluation of DELNA’s writing component. Read (2015) described correlations in 2001 for raters of DELNA conducted at the University of Melbourne’s LTRC ranging from an “unsatisfactory .79 to a highly acceptable .96” (Elder & Erlam, as cited in Read, 2015a, p. 207). Reliability analyses have not been routinely conducted; however, it was noted that

DELNA is double rated, with face-to-face training sessions held at the beginning of the semester for all raters, with additional sessions for new raters. An online rater training program was developed to complement the face-to-face rater training sessions and three empirical studies have been published presenting the efficacy of the online training (Elder et al., 2005; Elder, Barkhuizen et al., 2007; Knoch et al., 2007).

For PAAL, Knoch et al.'s (2016) application of the framework combined warrants 1, 2 and 5, to form their warrant 1. For the new first warrant, Knoch et al. (2016) reported the statistical properties of PAAL, referring to the test's trial, which involved 156 students from the University of South Australia ($n = 71$) and the University of Melbourne ($n = 85$) from a range of language backgrounds including students who spoke English as a first language (Elder & Knoch, 2009). Three additional, parallel versions of PAAL, for which statistical properties were presented (Knoch, 2010a, b, 2011, as cited in Knoch et al., 2016), were also referred to by the researchers. These studies by Elder and Knoch (2009) and Knoch (2010a, b, 2011) were commercial studies and not publicly available (U. Knoch, personal communication, February 13, 2018). Regarding the statistical properties of each version of the test, high to very high correlations ranging from .88 to .98 were noted (Knoch et al., 2016).

Warrants 3 and 4 focus on the administration procedures of a PELA. In particular, warrant 4 concerns the clear instructions, tasks, and as argued by Read (2015), the presentation to students of the task's rationale. Much of the PELA literature, including Read (2015), states that one potential problem area of PELAs is convincing EAL/D students to complete the assessment after they have demonstrated adequate ELP via proficiency tests, such as IELTS. Thus, it is imperative that the rationale for completing a PELA is convincing to achieve participation amongst this particular group of students. Another consideration worth noting is that English-speaking background students may question why they are completing a test, which results in judgements being made about students' use of their first language. Regarding DELNA, Read (2015) refers to a paper published in 2008 which outlines how DELNA staff promoted the PELA to students in a positive manner, thus satisfying the warrants.

For PAAL, warrant 3 (referred to as warrant 2 by Knoch et al., 2016) was supported based on questionnaire feedback from the initial trial (Elder & Knoch, 2009, as cited in Knoch et al., 2016) and the unpublished full trial of the test (Knoch & O'Hagan, 2014, as

cited in Knoch et al., 2016). A number of technical issues, such as the time it took for the webpage to load and lack of functionality across various platforms, were discovered initially. These issues were then rectified for the full trial. After the full trial, qualitative findings indicated “mostly positive” feedback regarding the online testing system (Knoch et al., 2016, p. 31). In comparison, the purpose of PAAL, including the anticipated benefits, were clear to test takers. Furthermore, the questionnaire results from the two trials demonstrated that students comprehended the instructions and requirements of the task (Knoch et al., 2016).

Finally, for warrant 5, Read (2015) addressed the difficulty level and item discrimination of the DELNA by referring to findings presented in two papers (Elder & Erlam, 2001; Elder & von Randow, 2008). The Screening test was found to be “easy for most candidates” (Read, 2015a, p. 206), yet demonstrated good item discrimination and acceptable reliability overall. The next section discusses the next inference in the framework, Generalisability.

3.7.2 GENERALISABILITY

The second inference, Generalisability, focuses on how generalisable a test’s forms and administrations are. In other words, “the assessment yields results that are consistent across assessment contexts” (Knoch & Elder, 2013, p. 56). Generalisability applies to PELAs that exist in multiple forms; such forms should be parallel regarding content and statistically equated so that test takers’ results are equal or close to equal, irrespective of which form of the test was taken (Read, 2015a). As noted by Read (2015a, p. 208), Generalisability “primarily involves the consistency of the assessment across different situations.”

The Generalisability inference comprises four warrants. The first warrant requires different forms of the test to be parallel in design. The second warrant concerns the use of appropriate equating procedures to ensure that different test forms entail equivalent difficulty. The third and fourth warrants necessitate the inclusion of sufficient tasks and consistent administration conditions (Knoch & Elder, 2013). Table 3.4 demonstrates the warrants and suggested supporting evidence to be obtained for the Generalisability inference.

Table 3.4*Warrants and potential supporting evidence for the Generalisability inference*

Generalisability inference	
Claim: The assessment yields results that are consistent across assessment contexts.	
Warrants	Sources of supporting evidence
1. Different test forms are parallel in design.	Review of test specifications and test materials
2. Appropriate equating procedures are used to ensure equivalent difficulty across forms.	Review of equating reports and statistical procedures used
3. Sufficient tasks are included to provide stable estimates of test taker ability.	Statistical analysis of scores from a trial test population
4. Test administration conditions are consistent.	Review of procedures; interviews with students and invigilators

Note: Adapted from Knoch and Elder (2013, p. 56). <https://doi.org/10.58379/YZLQ8816>. Permission granted from authors/copyright owners.

As displayed in the table, supporting evidence for warrants associated with the Generalisability inference includes review of test specifications and materials compared to extant literature, statistical analysis of pilot studies and review of procedures by key stakeholders, including students as test takers. The Generalisability inference is supported when observed scores are demonstrated to be estimates of expected scores in parallel versions of tasks (Knoch & Chapelle, 2017). The following section presents how the Generalisability inference of the PELA validation framework was applied to DELNA and PAAL.

This section now presents supporting evidence of each warrant for the Generalisability inference of DELNA and PAAL. Regarding DELNA, Read (2015) noted that as the PELA has grown in terms of test taker numbers and the number of institutions licensing the test, it was imperative DELNA designers ensured consistency across additional DELNA forms. In terms of warrants 1 and 2, new forms for the cloze-elide and writing components were developed within the DELNA Office. The listening and reading subtests are developed jointly by the University of Melbourne's LTRC and the DELNA Office. Trials comprising approximately 100 students occur for each new test with a comparison

undertaken against an existing form using the same student participants. A type of mathematical modelling known as Rasch analysis is subsequently performed in order to determine if revision is needed. The test is usually then trialled again at the University of Melbourne after revisions have been made (Read, 2015a).

In comparison, PAAL currently has four versions and are considered parallel (Knoch et al., 2016). Each version is based on the same specifications document regarding the first warrant. Knoch et al. (2016) referenced the publicly unavailable psychometric properties of versions 2, 3 and 4 (Knoch, 2010a, b, 2011, as cited in Knoch et al., 2016) that are said to closely resemble version 1 (Elder & Knoch, 2009). For warrant 2, the three variations of the test and the subsequent research detailing the statistical equating methods demonstrate support. No such data is available for the writing component. However, “the developers of the writing task attempt to closely stay true to the test specifications” with small trials of new writing subtest forms being evaluated by a team of developers to ensure equivalency and that they are eliciting samples of test taker writing that are assessable (Knoch et al., 2016, p. 33).

Concerning DELNA, reliability of new forms is routinely estimated with a minimum of .80 required responding to warrant 3 (Read, 2015a). Read also referred to test administration condition evidence conveyed under the Evaluation inference in support of warrant 4. For PAAL, Knoch et al. (2016) referred to the number of items included in the cloze-elide (75 items) and the c-test (75 items), as well as the writing task in response to warrant 3. The researchers noted that it is “practically impossible to add any more tasks” (Knoch et al., 2016, p. 33), as the duration of the test is already 60 minutes, considered at the upper limit of acceptable administration time. The researchers referred to reliability as reported in the trial studies. For warrant 4, the researchers noted, “as students can take the test in their own time at a place of their choosing, it is likely that the conditions are not absolutely consistent” (Knoch et al., 2016, p. 33). Some may complete the task in environments not conducive to concentration (e.g., a noisy, on campus computer laboratory), while others may complete it in optimal conditions away from distractions. Nevertheless, as PAAL is a low stakes test, the conditions the test are completed in “are probably not of great concern” (Knoch et al., 2016, p. 33). The next section explains the Explanation and Extrapolation inference.

3.7.3 EXPLANATION AND EXTRAPOLATION

The third inference in the validity framework is the Explanation and Extrapolation inference. The claim at this level of the validity argument is that “the assessment provides information on test takers’ skills/knowledge and characteristics that is in keeping with understanding of academic language proficiency and relevant to the academic domain. The test tasks are adequate proxies for those performed in the academic domain” (Knoch & Elder, 2013, p. 57). This inference focuses on how the tasks of a test and the elicited language compare to task types and language utilised in the academic domain. It also includes the accuracy to which assessment tasks predict language performance in academia or in other assessments that have been designed to assess academic English (Knoch & Elder, 2013). Explanation refers to the relationship between the observed test performance and the construct being assessed (Chapelle et al., 2008). In contrast, extrapolation refers to the inference made when a test taker’s expected score, defined as the score a student is likely to receive on similar test tasks, indicates performance in the target setting (e.g., a university subject). Combined, Explanation and Extrapolation involves investigating whether the PELA is relevant and appropriate within the domain of academia (Knoch & Elder, 2016).

The complex nature of the Extrapolation and Explanation inference requires six warrants. The first warrant examines whether the assessment results are adequate predictors of language performance in academia. The second requires similarity between the PELA tasks and tasks required by students at university. The third states that test takers’ linguistic knowledge, processes and strategies utilised when completing the assessment are consistent with what is required within academia. The fourth and fifth warrants maintain that PELA scores provide information about test takers’ academic language proficiency and are related to performance on alternative assessment measuring the same construct. Importantly, the sixth warrant ensures that PELA tasks are fair for all test takers (Knoch & Elder, 2013). The warrants and suggested supporting evidence to be obtained for the Explanation and Extrapolation inference as detailed in Table 3.5.

Table 3.5*Warrants and potential supporting evidence for the Explanation and Extrapolation inference***Explanation and Extrapolation inference**

Claim: The assessment provides information on test takers skills/knowledge and characteristics that is in keeping with understanding of academic language proficiency and relevant to the academic domain. The test tasks are adequate proxies for those performed in the academic domain.

Warrants	Sources of supporting evidence
1. Test results are good predictors of language performance in academic domain.	Correlations between PELA scores and academic performance esp. language-related academic tasks (e.g., essays, oral presentations, GPA and weighted-average marks [WAM])
2. Characteristics of test tasks are similar to those required of students in the academic domain (and those in the language development courses students are placed in).	Comparison of test materials and course materials/course assessment requirements
3. Linguistic knowledge, processes, and strategies employed by test takers are in line with theoretically informed expectations and observations of what is required in the corresponding academic context.	Test taker verbal protocols gathered during test performance; responses on strategy questionnaires gathered after the test
4. Scores derived from the test provide sufficient information about candidates' academic language proficiency (i.e., no construct under-representation).	Review of test materials to ensure adequate coverage of academic language domain
5. Performance on PELA relates to performance on other assessments of academic language proficiency.	Correlation between test scores and scores derived from other validated test instruments measuring similar abilities;

Explanation and Extrapolation inference

Claim: The assessment provides information on test takers skills/knowledge and characteristics that is in keeping with understanding of academic language proficiency and relevant to the academic domain. The test tasks are adequate proxies for those performed in the academic domain.

	correlation between test scores and teacher rankings of language proficiency
6. Tasks do not unfairly favour certain groups of test takers.	Expert ‘sensitivity’ review of test content, statistical bias analyses

Note: Adapted from Knoch and Elder (2013, pp. 57-58). <https://doi.org/10.58379/YZLQ8816>. Permission granted from authors/copyright owners.

The table above indicates that support for warrants attached to the Explanation and Extrapolation can be found via correlational analysis. Such analysis may include comparing PELA scores and scores on other university assessments that examine academic language proficiency and other measures of language proficiency (e.g., IELTS). Furthermore, literature review of tasks in the academic domain and feedback from stakeholders, especially test takers, regarding their strategies and opinions on test fairness comprise possible sources of evidence. Investigations at this level can also involve “comparing assessment scores to criterion scores based on a thorough (and representative) sample of performances from the target domain” (Kane et al., 1999, p. 10). Support for inference is shown when test performance reflects the language requirements of the target language use domain of interest; hence, the rating procedures of a PELA should model rating procedures of language performance in the target language use domain (Knoch & Chapelle, 2017). The following section presents how the Explanation and Extrapolation inference was applied to DELNA and PAAL.

This section now presents evidence that comprised support for the Explanation and Extrapolation inference of DELNA and PAAL.¹ Warrant 1 involves determining whether results on the test are indicative of students’ academic performance. For DELNA, correlations were calculated between DELNA scores and GPA in the first two semesters of

¹ Knoch et al. (2016) presented the warrants in a different order to the original framework. The order was rearranged by the researcher for consistency with Knoch and Elder (2013) and Read (2015)

university as reported in Elder, Bright et al. (2007). Although there were variations amongst faculties and assessed skills (i.e., writing, reading, or listening), overall correlations ranged from .30 to .35 over the four-year study, consistent with the findings in other studies focusing on large-scale ELP tests (Read, 2015a). A further source of evidence was the academic outcomes of those who were recommended to proceed to DELNA Diagnosis but elected not to, referred to as “avoiders” (Read, 2015a). Concerning PAAL, no data was collected responding to this warrant. However, an internal report determined the correlation between the DELA, which was said to have a strong relationship with PAAL (Knoch et al., 2016), and test takers’ weighted average marks (WAM). WAM is similar to GPA, but is an average of students’ actual marks, not grades. The report demonstrated that “a higher score in DELA is associated with higher WAMs and that a higher DELA score is associated with lower risk of failing” (Knoch et al., 2016, p. 35).

Warrants 2 and 4 concern the relationship between the assessment and the tasks which students will complete in their degree programme. For warrant 2, Read (2015) noted that an opportunity to review the reading and listening components arose in 2012 when it was decided to require first year, Higher Degree Research (HDR) students to complete DELNA. Read (2015, p. 212) argued that “warrant 4 is rather broadly stated”, further noting that the desired “adequate coverage of the academic domain” (Knoch & Elder, 2013, p. 57) is difficult to determine. Read (2015, p. 212) also argued that fully representing each construct is “an unattainable goal for any assessment.” This emphasised the difficulty in obtaining evidence for certain warrants and, therefore, inferences that form the validity argument for a PELA. Hence, flexibility is required. Turning to PAAL, no data was collected for warrant 2. For warrant 4, Knoch et al. (2016) stated that the rubric for the writing subtest was developed based on the English for academic purposes experience of developers and upon review of writing assessment in academia. Knoch et al. (2016, p. 34) emphasised that “the criteria used on the scale (organization and style, content and form) are commonly used in the assessment of academic writing and the level descriptors have been refined over the years to assist raters in better differentiating between candidates.” It was noted that, as the time limit for the written task is 30 minutes, “only a limited sample” of test takers’ written abilities is elicited (Knoch et al., 2016, p. 34). Consequently, the researchers posited that the rubric is limited; however, it measures writing skills representative of the academic domain to a certain extent.

Warrant 5 mostly concerns concurrent validation, involving the correlation of one, new test with another well-established instrument, measuring the same construct. Read (2015) explained the decision was made not to use high stakes language test scores, such as IELTS, to correlate with DELNA, as the PELA is administered to English-speaking background students, who are not required to complete a language test, such as IELTS, pre-entry, as well as EAL/D students. In contrast, the University of Auckland chose to use grades in academic subjects as the criterion measure to validate DELNA scores (Smith, as cited in Elder, Bright et al., 2007). Read (2015) suggested an alternate method may be to obtain academics' views of the written piece and how they would assess it. Regarding PAAL, Knoch et al. (2016) reported the direct comparisons made between performance on PAAL and DELA by students who completed both PELAs. Results of the comparison revealed significant, moderate to high correlations were demonstrated from .699 for the cloze-elide to .809 for the screening total scores.

Finally, warrant 6 looks at the fairness of the task. Notably, this warrant is seen by Read (2015, p. 213) as “potentially of concern to DELNA” due to the inclusion of both English-speaking background and EAL/D students completing the assessment. Elder and Erlam (2001) determined bias to be an issue when the test was first employed. Subsequently, Elder et al. (2003) presented a methodology for the investigation of bias utilising Rasch analysis, and, although the study included a small sample size, a statistically significant differential item functioning was found for the two groups of students in all components of the DELNA, except the vocabulary test.

For both DELNA and PAAL, evidence to support certain warrants was not obtained by the researchers. Warrant 3, which requires test takers' verbal reports of their cognitive perspectives of the task to determine whether they apply expected skills in completing the task, was not collected regarding DELNA nor PAAL. Furthermore, evidence was not presented in response to warrants 2 or 6 concerning PAAL. Knoch et al. (2016) suggested that the warrants for which evidence was available were supported to a satisfactory level. However, it was noted that due to limited scope and the PELA's screening function, there is reduced capacity for PAAL to be fully representative of the academic language domain (Knoch et al., 2016). The next section analyses the Decisions inference of Knoch and Elder's (2013) framework.

3.7.4 DECISIONS

The fourth inference, Decisions, focuses on results and recommendations after the assessment has been completed and scores communicated. Here, the claim is that “score-based decisions are appropriate and well communicated” (Knoch & Elder, 2013, p. 59). It includes the accuracy of student categorisations, based on academic language proficiency, described by the researchers as often neglected. Other considerations at this level include the timeliness of results and feedback, how closely linked recommendations are to on-campus support, and who has access to results and recommendations. Overall, the decisions inference connects test takers’ scores to decisions made about individuals upon scores (Chapelle et al., 2008).

The Decisions inference is made up of six warrants. The first and second warrants require students to be categorised accurately in terms of academic language proficiency, as well as the inclusion of feedback and recommendations. The third warrant necessitates the recommendation being closely linked to university support services, such as ALL units. The fourth and fifth warrants ensure timely distribution of results and availability of such results for all stakeholders. Finally, the sixth warrant determines that test users understand both the meaning and the intentions of PELA scores. Table 3.6 details the warrants and suggested supporting evidence to be obtained for the Decisions inference.

Table 3.6

Warrants and potential supporting evidence for the Decisions inference

Decisions inference	
Claim: Score-based decisions are appropriate and well communicated.	
Warrants	Sources of supporting evidence
1. Students are correctly categorised based on their test scores.	Interviews with key stakeholders (e.g., students, academic staff, learning and teaching staff); review of test results; review of standard-setting activities to set cut-scores; review of academic outcomes for students classified above and below the cut-score

Decisions inference

Claim: Score-based decisions are appropriate and well communicated.

2. The test results include feedback on test performance and a recommendation.	Review of policy and practice
3. Recommendation is closely linked to on-campus support.	Review of language development options; interviews with key stakeholders including students
4. Assessment results are distributed in a timely manner.	Review of practice; interviews with key stakeholders
5. The test results are available to all relevant stakeholders.	Review of policy and practice
6. Test users understand the meaning and intended use of the scores.	Review of policy and practices, including test website; interviews with test users

Note: Adapted from Knoch and Elder (2013, p. 59). <https://doi.org/10.58379/YZLQ8816>. Permission granted from authors/copyright owners.

As displayed in the table, the Decisions inference requires “evidence that the assessment results provide an appropriate basis for advising students on whether they need to enhance their academic language ability and, if so, how” (Read, 2015a, p. 213). Therefore, it involves determination of whether stakeholders at the university who utilise the results are aware of how they are correctly interpreted. Consequently, evidence supporting the six warrants may include consultations with stakeholders and comparative analyses of performance on other validated measures. Insight can also be sought from stakeholders on aspects such as timeliness of the distribution of results and feedback and how closely linked recommendations are to on-campus support mechanisms (Knoch & Elder, 2013). This process may include consultations with stakeholders and comparative analyses of performance on other validated measures. Support is formed for the inference when decisions made, based on the test’s results, are appropriate and effectively conveyed to test stakeholders (Knoch & Chapelle, 2017). Knoch and Elder (2016) noted that for the university context, the final two inferences, Decisions and Consequences, which transition from the interpretation of scores’ meaning to scores’ utilisation (Chapelle et al., 2008), are of particular importance, as any decision made based on PELA scores is crucial in achieving the

goal of enhancing student retention and success. The following section presents how the Decisions inference of the framework was applied to DELNA and PAAL.

This section presents evidence backing the Decisions inference of DELNA and PAAL. The first warrant regards the correct categorisation of students based on PELA scores. Regarding DELNA, Elder and von Randow (2008) presented their rationale and processes for determining cut scores for DELNA's Screening test. A cut score, which results in test takers proceeding to the Diagnostic, that was sufficiently sensitive to identify 93% of students at risk of academic failure due to low English language proficiency was determined (Elder & von Randow, 2008). However, the cut score eventually applied was set "relatively low" (Read, 2015a, p. 212). This was due to budgetary concerns; however, students scoring above the cut score receive "generic advice on how to enhance their academic language skills" (Read, 2015a, p. 212). For PAAL, gathering support for this warrant entailed two standard-setting processes. The first was during the development of the test, via a Receiver Operating Characteristics curve analysis, a technique for standard setting, to determine cut-scores on the Screening test. The second process involved a team of trained language assessors setting the cut-scores of the writing subtest through individually rating and then discussing their ratings of 50 writing texts (Knoch et al., 2016).

Warrant 2 requires that results provided to students contain feedback and a recommendation. In the context of DELNA, the most detailed feedback is provided to DELNA candidates who are identified as most at-risk of academic failure during a 30-minute consultation with a Language Advisor at the DELNA Office, during which results are reviewed (Read, 2015a). In contrast, feedback provided to students for PAAL is minimal. This was reported by student participants of the full trial, who stated they were disappointed with the statement of results whilst calling for more detailed diagnostic analysis of their performance face to face with an advisor (Knoch et al., 2016). Participants in the full trial had issues with the "vagueness of the support recommendation given in the report, with many of them wanting a clearer directive for what was required of them" (Knoch et al., 2016, p. 37).

Warrant 3 states that recommendations provided to students based on PELA scores are linked to on-campus support. For DELNA, recommendations for utilising on-campus support are made, thus supporting warrant 3 (Read, 2015a). For PAAL, feedback on the full trial indicated concerns regarding availability of support mechanisms (Knoch et al., 2016). It was felt that support was not suited to students' academic language proficiency, programme

level, academic field, or their needs. There was also concern regarding a perception that such support mechanisms would incur further costs.

Warrant 4 concerns the timely reporting of results. For DELNA, results of the Screening test are automated and sent to students within 24 hours. For the Diagnostic test, due to the writing component being cross-marked, there is a 10-day turnaround for students (Read, 2015a). In PAAL's context, results are distributed within one to two days after the assessment has been taken. Students appreciated this feature as noted in the full trial (Knoch et al., 2016).

Warrant 5 relates to the availability of PELA results. Read (2015, p. 214) suggested partial support for DELNA, maintaining: "results are also provided on request to the relevant stakeholders." In comparison, university stakeholders can request PAAL results from the LTRC at the University of Melbourne. However, it was noted by the researchers that academic staff, including lecturers, may not be aware they can access PAAL results (Knoch et al., 2016).

The final warrant requires understanding on the part of test takers concerning the meaning and intended uses of PELA scores. Regarding DELNA, no investigation had been conducted to gauge test users' understanding of the interpretation of DELNA results. However, it was explained that this is done face-to-face by the DELNA Manager regularly due to the turnover of both academic and professional staff. A four-page "frequently asked questions" guide is also distributed to various stakeholders across the university to provide information about DELNA and the process of support (Read, 2015a). Additionally, a DELNA Reference Group meets twice per year to discuss issues and inform stakeholders of developments in the DELNA procedure. In contrast, the full trial evaluation of PAAL indicated students understood the intended purpose (Knoch et al., 2016). However, this commentary referred to student stakeholders, not others who may use and make decisions based on results, thus, further investigation is necessary to fully support the warrant. The next section presents the final inference in the framework, Consequences.

3.7.5 CONSEQUENCES

The final inference, the Consequences inference, concerns the claim that "the consequences of using the PELA and the decisions informed by the PELA are beneficial to all stakeholders" (Knoch & Elder, 2013, p. 60). Hence, the Consequences inference necessitates

evidence of beneficial outcomes from the PELA for students especially, as well as for the institution overall (Read, 2015a). Like the Decisions inference, the PELA policy of the university and overall context in which it is integrated governs the Consequences inference. Perhaps the most obvious consequence is that students who have been identified as requiring academic language development will, in fact, participate in the process of support.

The Consequences inference is the most complex element of the framework, composed of eight warrants. The first warrant proposes that all targeted test takers complete the assessment. The second and third warrants preclude stigma and disadvantage for students and require positive perceptions of the assessment amongst test takers. The fourth and fifth warrants necessitate feedback informing learning and students acting on any recommendation made. The sixth and seventh warrants ensure language development opportunities post-PELA are appropriate, and learners utilising support develop their academic language skills. Finally, the eighth warrant predicts that students who do not respond to recommendations perform less successfully than their counterparts who do (Knoch & Elder, 2013). Table 3.7 presents warrants and potential support for the Consequences inference.

Table 3.7

Warrants and potential supporting evidence for the Consequences inference

Consequences inference	
Claim: The consequences of using the PELA and the decisions informed by the PELA are beneficial to all stakeholders.	
Warrants	Sources of supporting evidence
1. All targeted test takers sit for the test.	Analysis of test administration data
2. The test does not result in any stigma or disadvantage for students.	Interviews with students regarding their test attitudes and experiences following the test
3. Test takers' perceptions of the test and its usefulness are positive.	Interviews with students
4. The feedback from the test is useful and directly informs their future learning.	Interviews with students and language support teachers

Consequences inference

Claim: The consequences of using the PELA and the decisions informed by the PELA are beneficial to all stakeholders.

5. Students act on the test recommendation (i.e., take up the proposed language development strategies).	Review of student uptake data. Interviews with key staff and students regarding reasons for compliance or non-compliance
6. Follow-up language development options provided for students are appropriate.	Interviews with key stakeholders including students
7. Learners taking up support options improve their English over the course of their studies.	Comparison of pre- and post-test scores
8. Students who fail to act on test recommendations are more likely to struggle in their academic studies.	Comparison of academic results of compliant and non-compliant students

Note: Adapted from Knoch and Elder (2013, p. 60). <https://doi.org/10.58379/YZLQ8816>. Permission granted from authors/copyright owners.

As shown in the table, evidence gathered to support warrants attached to the Consequences inference include analysis of students' perceptions of a PELA including the feedback they receive, evidence to support the development of academic language (e.g., pre and post-test scores) and academic outcomes of those who participate in the PELA process and those who do not. Students' perceptions of a PELA including the feedback they receive requires exploration. This insight has appeared infrequently in PELA literature apart from Elder and von Randow's (2004) and Bright and von Randow's (2008) examination of students' perceptions of DELNA. Arguably the most difficult to capture is evidence supporting the development of ELP as a result of students participating in support mechanisms (Knoch & Elder, 2013). Regardless, "such evidence, however tentative, needs to be sought if claims about the benefits of any PELA initiative are to be upheld" (Knoch & Elder, 2013, p. 61). Overall, this inference necessitates "evidence of positive outcomes from administering the assessment for the students in particular as well as for the institution as a whole" (Read, 2015a, p. 215). The Consequences inference is justified when the

consequences of a test are aimed to promote the interests of test users, thus having positive consequences (Knoch & Chapelle, 2017). The following presents how the Consequences inference was applied to DELNA and PAAL.

This section summarises the evidence comprising support for the Consequences inference of DELNA and PAAL. The first warrant calls for all target test takers to take the test, which, as Read (2015) explained, is not easily achieved nor determined. It is estimated that “around 90 per cent” sit the Screening component while “around 60 per cent” complete the Diagnostic test as recommended (Read, 2015, p. 215). The streamlining of PAAL was so that it could be universally administered at the University of Melbourne (Knoch et al., 2016). However, it was apparent institutional policy made it impossible to make such an assessment compulsory. This was stated to be “because it goes beyond content course requirements” (Knoch et al., 2016, p. 38). Completion rates of students enrolled in two degrees, the Bachelor of Commerce and the Master of Engineering demonstrated only 35% (110 out of 310 students) and 12% (60 out of 491 students) completed the PELA respectively (Knoch et al., 2016). Furthermore, focus group data suggested students were unsure who the test was for, “with many believing it to be intended for ‘international’ students only” (Knoch et al., 2016, p. 38).

Warrant 2 concerns stigma associated with the students who have completed the test and/or based on their results. Read (2015, p. 218) argued that “the fact that university policy does not exempt any students from the assessment reduces, if not eliminates, any sense of stigma associated with taking the DELNA screening.” Knoch et al. (2016) did not respond to the second warrant in their investigation of PAAL.

The third warrant states that the perception amongst test takers is positive. In the case of DELNA, warrants 3 and 4, which call for feedback from students, are seen by Read (2015), as being met by procedurally monitoring students’ opinions and DELNA experiences via an online survey and interviews. The survey asks students to comment on DELNA, how their English skills have affected their university experience and any language support accessed. Students who respond are also able to participate in an interview to discuss their responses in more detail. Feedback from students resulted in changes made to DELNA (Read, 2015a). For PAAL, the finding from the full trial determined mixed feelings. Students were positive towards the relative ease of completing the task online yet were less positive in

terms of the feedback afforded to them and the support options available to them (Knoch et al., 2016).

Warrant 4 requires feedback to be useful and directly inform student learning. Read (2015) commented on warrants 3 and 4 jointly, as discussed above. In the case of PAAL, feedback from the full trial indicated students did not find the feedback especially useful. Nevertheless, as argued by Knoch et al. (2016, p. 39), “PAAL is intended as a screening test, which is designed to identify students deemed to be at risk in minimum time and with minimum financial expenditure”; thus, resources to deliver detailed feedback were not available. The researchers concluded here that “unfortunately, offering more varied support provisions is costly and, in the current climate of cost-savings, probably not a viable option in the near future” (Knoch et al., 2016, p. 39).

The fifth warrant states that test takers act on the recommendation delivered. It, therefore, requires determination of whether students act on recommendations based on PELA results. For DELNA, a qualitative study by Bright and von Randow (2004, p. 4) found that although each student had taken up a language enhancement recommendation, “it was in many cases sporadic and unfocused in nature and unlikely to address their real needs in terms of academic language proficiency.” Thirteen of the 18 students who participated in the study were re-interviewed after their graduation (Bright & von Randow, 2008). It was, subsequently, found that the students did improve their academic literacies. Read (2015) also used this evidence in response to Warrant 8, which predicts students who fail to act on test recommendations are more likely to struggle at university. For PAAL, approximately 15% of students were categorised as being at risk of academic failure. However, Knoch et al. (2016) stated that it was unclear how many of these students acted on the recommendation of enrolling in language enhancement courses. Historical enrolments in such courses were stated to be low, expressing a lack of support for this warrant.

Warrant 6 examines the appropriateness of the language development recommendations. An English Language Support Taskforce was established, which presented recommendations to the University of Auckland concerning language enhancement policy, yet additional resources were not committed as per the recommendation (Read, 2015a). No systematic review of student services has been conducted. Warrant 7 seeks to determine if language proficiency is improved. No data for DELNA was presented here. For PAAL, warrants 6 through 8 were not investigated in the study by Knoch et al. (2016).

Overall, the review of the two PELA contexts has demonstrated that DELNA and PAAL have a variety of strengths; however, support for certain warrants and, therefore, inferences was not necessarily determined. These areas include features of the assessment itself and the university policies in which the PELAs are placed. Regarding DELNA, evidence was demonstrated to support most of the warrants for each of the five inferences. However, Read (2015) noted that “much of the evidence is partial and somewhat informal in nature” (p. 219). Read (2015, p. 219) acknowledged that the validation study was conducted by those involved with developing, implementing and/or monitoring the respective PELAs; thus, the investigation “represents an insider’s view of the assessment programme”. Importantly, the validation efforts have allowed identification of the areas which necessitate attention and stakeholders to come to their own conclusions concerning validity.

Turning to PAAL, Knoch et al. (2016) presented a variety of evidence types from trials to support the warrants of the validity framework. Warrants backing the first three inferences of the framework, Evaluation, Generalisability and Explanation and Extrapolation were supported. Similar to DELNA, backing for warrants connected to the final two inferences presented a “mixed picture” (Knoch et al., 2016, p. 40). Thus, the inferences most affected by university policy, the Decisions and Consequences inferences, are the two that require further investigation. As surmised by Knoch et al. (2016, p. 41), although PAAL’s design and the information it offers regarding students’ needs is sound, “there are issues with its utilization that require attention. Most of these issues are related to the policy environment ... rather than the nature of the test itself.” This reveals an important consideration, as the policy environment is not an external constraint, but an important aspect of validity.

3.8 CONCLUSION

This chapter focused on PELAs and validity. PELAs are administered to determine academic language needs of students already admitted into university programmes. The justifications of PELAs are clear; firstly, they make students aware of the importance of effective academic language and communication skills. PELAs identify those students at potential risk of academic failure. Support must then be provided in order for such students to develop in the areas requiring it. Meanwhile, PELAs lead to the promotion of academic language and learning units; they are often the conversation starter between students and

advisors. Finally, PELAs demonstrate that universities are serious about assisting students in the development of communication skills, commonly seen in university graduate attributes.

Although the overall intent of PELAs is clear, there are several nuances to different universities' post-admission assessments. No two PELAs are identical. This chapter has analysed the salient features, such as targets, mandatory completion, skills assessed and links to on-campus support, with reference to extant literature and available case studies. Numerous considerations to be taken into account were examined. The first is that of practicality. PELAs can be resource intensive, from design to communicating results and providing support. The next consideration discussed is difficulty in getting to students who most need support. PELAs and related support mechanisms struggle to reach ideal participation and engagement goals. It may be relatively easy to require students to complete an assessment; it is much harder to get them to take up the offer of support, particularly when they are already studying full-time. The most important consideration, however, is that of validity. To date only few studies have been published presenting validity arguments for PELAs. If PELAs are to be taken seriously by stakeholders, more must be done in this regard.

Turning to language assessment validity, conceptual understanding and practice have evolved considerably. Validity used to concern whether an assessment measures what it claims to measure. Contemporary understanding posits a more holistic, unitary evaluation of the extent to which evidence and theory support the appropriacy of interpretations and decisions made based on assessment scores. Messick's (1989) proposal of an integrated view of validity sees assessment as a reasoning and evidence gathering process carried out to make inferences about test takers. Hence, the validation process requires the gathering of a broad range of data that adds to the evaluation of an assessment's validity overall, known as the presentation of a validity argument (Davies & Elder, 2005). Therefore, validation frameworks have played an integral role in guiding researchers in the process of collecting, collating and presenting validity evidence.

An important framework is the argument-based approach to validation, which was developed to provide clarity and strategies, in a pragmatic way, to allow sufficient flexibility for different contexts (Kane, 2012). As per the argument-based approach to validation, for any inference, a claim is put forward. Warrants, described as assumptions that underly the stated claims, are then formulated. Each claim contains multiple warrants depending on the assessment context and warrants must be supported by empirical and/or theoretical evidence.

The process of presenting a validity argument, thus, requires the gathering of a broad range of data thereby going through the inferences, claims and warrants at each level of the validity argument (Davies & Elder, 2005). The validity argument that is presented for one language assessment may be quite different to the argument presented for another.

Based on the literature discussed, Knoch and Elder (2013) put forth a validation framework encompassing an interpretive argument to assist in studies of PELAs specifically. The framework is a hybrid model of the works of Kane (1992) and Kane et al. (1999), as well as Bachman and Palmer's (2010) assessment use argument and includes a series of inferences for which validity evidence is required. The five inferences that form the framework include: Evaluation, Generalisability, Explanation and Extrapolation, Decisions and Consequences. In addition to the five inferences and their associated claims, warrants are attached to each inference. These warrants are based on language assessment literature and studies that have applied the argument-based approach to validation. Overall, the framework assists in making determinations surrounding a PELA's ability to aid appropriate decision making as to which students may benefit from language skills development and whether the PELA process makes a contribution to student achievement (Read, 2019). The validity framework is appropriate for use in this study, as it is grounded on seminal validity and validation literature.

The presentation of the validity framework applied to two university PELAs, DELNA and PAAL, demonstrates that it allows for an adaptable framework for investigating validity. The sources of evidence to be used when applying the framework can be categorised into three broad categories: reviewing current practices based on best practice in the literature, obtaining qualitative data via interviews, and conducting quantitative research by performing statistical analyses on test data (e.g., scores, properties) and determining correlations (e.g., PELA test score and students' GPA). Knoch and Elder (2013) noted that the list of warrants is not exhaustive, nor is data required for each warrant. Read (2015) also posited that as PELAs are low stakes assessments, the amount of validity evidence required is lower than that of a commercial, standardised test. The review of the two PELA contexts has demonstrated that each PELA has a variety of strengths; however, support for certain warrants and, therefore, inferences was not necessarily determined. These areas include features of the assessment itself and the university policies in which the PELAs are placed. This issue applies to many PELA contexts; thus, "those charged with implementing such

initiatives should perhaps temper their claims about what their assessments are for and what they can achieve” (Knoch & Elder, 2016, pp. 227 – 228).

This thesis aligns with the view that a validity argument supporting both interpretation and use of test scores ought to be based on an “overall interpretive argument” (Chapelle et al., 2008, p. 6). Much like the applications of the framework that were discussed in this chapter, this thesis sought to construct a sound and coherent argument in support of one such PELA. Chapter 5 presents the methodology undertaken in the current research to present a validity argument for the PELA and determine what factors, if any, play a role in academic performance within the context of one Australian university.

CHAPTER 4: METHODOLOGY

4.0 INTRODUCTION

This chapter describes the methodology of the thesis. First, it presents the research aims and research questions. The chapter then explains the study's overall research design, followed by the sample and sampling strategies employed. Ethical considerations are then detailed. The chapter subsequently presents the research tools utilised in the study, as well as the analysis procedures used. Finally, the chapter elucidates how the research questions were answered. This chapter concludes with a summary of the methodology.

4.1 AIMS AND RESEARCH QUESTIONS

This thesis had two research aims. The overarching aim was to determine possible factors able to predict the academic performance of undergraduate students at tertiary level. For the purposes of this study, academic performance was defined as results obtained by students including overall grade outcomes and GPA. The examined factors included undergraduate students' academic essay writing, as measured by a PELA, Bond English Language Assessment (BELA), and academic performance at university, over two semesters of study, as well as non-academic factors, including demographic data. It was envisaged that the analyses and subsequent findings would lead to the development of a profile of undergraduate students who may require additional academic support. This would subsequently assist stakeholders at the University, in providing targeted support to students who most require it to contribute to the University's retention strategy and assist students in achieving academic success. This research may also assist other universities to identify students who are potentially at risk of academic failure or not achieving their academic goals, thus contributing to enhanced retention and success rates for undergraduate students at an early stage in their studies.

A secondary aim was to present a validity argument for BELA, a post-entry measure of academic essay writing skills of university students enrolled in a compulsory undergraduate subject, Core 1: Critical Thinking and Communication. Knoch and Elder's (2013) validity framework, outlined in Chapter 3, was employed to achieve this aim. Presenting a validity argument for BELA was necessary to achieve the overarching aim of predicting academic performance, as BELA is a key variable included in the current investigation.

Two main research questions were formulated. Research question 1 concerned the validity argument for BELA and comprised an overarching question with five sub-questions. The overarching question was:

RQ1: Can a satisfactory validity argument be presented for the Bond English Language Assessment (BELA)?

In order to respond to the first research question, Knoch and Elder's (2013) framework for PELA validation was utilised. Accordingly, the first research question contained five sub-questions, adapted from inferences in the framework. The applicable inference is noted in parentheses after each sub-question, as follows:

RQ1a: Are BELA scores adequate reflections of observed behaviours? (Evaluation inference)

RQ1b: Does BELA yield results consistent across assessment contexts? (Generalisability inference)

RQ1c: Does BELA provide information on test takers' skills, knowledge and characteristics that keep within the understanding of academic English language proficiency? Is BELA an adequate proxy for tasks performed in the academic domain? (Explanation and Extrapolation inference)

RQ1d: Are decisions based on BELA scores appropriate and well communicated? (Decision inference)

RQ1e: Are the consequences of using BELA and the decisions informed by BELA beneficial to all stakeholders? (Consequences inference)

The second research question presented below referred to the overarching aim of the thesis, factors which may predict academic performance of students at the University.

RQ2: What factors predict academic performance of university students?

In summary, this study sought to determine whether BELA results and/or additional factors, such as early academic performance, age, gender, and language background, were able to adequately predict the academic performance of undergraduate students at university

after two semesters of study. It was, first of all, necessary to ascertain whether a satisfactory validity argument could be presented for BELA.

4.2 OVERALL RESEARCH DESIGN

This study employed a mixed methods approach (Creswell & Plano-Clark, 2018; Dörnyei, 2007; Riazi & Candlin, 2014), focusing on quantitative data, whilst including qualitative data to gain broader understanding of the research matter (Creswell & Guetterman, 2019). The current study utilised the mixed methods approach as this may increase the research validity of studies via “the convergence and corroboration of the findings” (Dörnyei, 2007, p. 45) and is considered effective for its provision of “the most informative, complete, balanced, and useful research results” (Johnson et al., 2007, p. 129). A mixed methods approach was particularly advantageous in investigating a sound validity argument for BELA (i.e., RQ1), as it assisted in providing systematic and comprehensive analysis of stakeholders’ views. The next sections of this chapter detail the quantitative and qualitative data obtained in the current investigation.

4.2.1 QUANTITATIVE DATA ANALYSIS

Quantitative data analysis involves quantifying certain aspects of language development or factors that may influence language learning and use (Phakiti, 2010). In the current study, quantitative data was collected and analysed in four categories: 1) student background, 2) academic performance, 3) questionnaires, and 4) BELA reliability data. This section describes these four categories.

The first category, student background data, included demographic information about the students enrolled in the subject, Core 1: Critical Thinking and Communication, at the University. Demographic data comprised the following: age, citizenship, country of birth, gender, language background (i.e., main language spoken at home), identification as Aboriginal or Torres Strait Islander, and identification as being faced with a disabling condition.

Information pertaining to academic performance formed the second category of data. This information included BELA scores, Major Essay marks, and overall grade in Core 1: Critical Thinking and Communication. In addition, students’ GPA after completing two semesters of university was utilised as students’ academic achievements at this point of their

studies may indicate their performance across the entirety of their study programmes (Gershenfeld et al., 2015; Humphreys et al., 2012).

Three questionnaires provided the data for the third category of quantitative analysis, with the opinions of both students and staff forming integral sources of evidence (Read, 2010). Consequently, the following three groups of stakeholders completed a targeted questionnaire: students enrolled in Core 1: Critical Thinking and Communication, academic staff teaching on the subject, and other users who utilised BELA scores. Quantitative items in the three questionnaires obtained data concerning attitudes toward BELA, including its usefulness, administration procedures, test specifications and materials, policy and practice regarding assessment of academic writing skills. Section 4.5.3 of this chapter includes specific details about the questionnaires.

The final category of quantitative data, BELA reliability, was addressed through analysing inter-rater reliability. This procedure, employed to minimise rater subjectivity in assessing student writing (Roever & Phakiti, 2018), involved comparative analysis of the original BELA ratings by academics and the researcher (the second rater), and, additionally, by two external raters. This was necessary as BELA's efficacy depends on the accuracy of scores to diagnose academic writing needs of students. Additional ratings were, thus, carried out by the researcher and the two external raters to provide a comprehensive evaluation of inter-rater reliability. The researcher, firstly, rated randomly selected BELA essays of 180 students of the 1,116 total completions (16.1%), to compare ratings with the original ratings of academics. Two additional raters (i.e., External rater 1 and External rater 2) were employed to assess 120 BELA scripts (10.8%) to ensure unbiased review of BELA.

4.2.2 QUALITATIVE DATA ANALYSIS

The qualitative data was obtained and analysed to complement quantitative analysis and “to get to the bottom of what is going on in all aspects of social behaviour” (Holliday, 2010, p. 99). Consequently, qualitative analysis provided meaningful insight into stakeholders' interpretations and perceptions of BELA. Additionally, two external raters provided feedback via semi-structured interviews, allowing objective, unbiased evaluation of BELA. Qualitative data were collected and analysed in three categories: 1) questionnaires, 2) Academic Skills Centre feedback surveys, and 3) semi-structured interviews.

The first category was the three questionnaires introduced in section 4.2.1, which included open-ended items providing an opportunity for respondents to add further information in addition to their responses to quantitative items. Open-ended questions allowed stakeholders to provide information concerning strategies utilised during the assessment, efficacy of language development opportunities and feedback provided to students after completion of BELA. This qualitative analysis complemented objective quantitative procedures (Read, 2010).

Qualitative data was also obtained via the Academic Skills Centre (ASC) online feedback survey sent to students who had attended at least one consultation. The Survey Monkey questionnaire, sent yearly, includes the following open-ended question relating to students' experiences with BELA and follow up support provided by Academic Skills Centre: "Did you do the BELA writing task for Core 1 this semester?". The respondents indicating "Yes" were prompted to "Tell us more about your experience" In total, 52 student responses were gathered from the second item in the feedback survey.

Finally, qualitative data was obtained through two 30 min Microsoft Teams semi-structured interviews with each of the two external raters, experienced language assessors, employed to rate approximately 10% of the BELA scripts. Upon completion of rating moderation, the researcher conducted an initial 30-minute semi-structured interview collecting data via open-ended questions regarding BELA, rating BELA and the procedure of identifying students who may require additional academic writing support (see Appendix E for indicative interview questions). The aim of the interviews was to obtain initial feedback on BELA and its associated processes. Two final interviews were then conducted once the external raters had completed their ratings. The interviews with the external raters provided information on the assessment itself, including its usefulness, appropriateness, specifications and materials, as well as administration procedures, policy and support provided to students.

4.2.3 RESEARCH SITE

The research site for this research project was Bond University, Gold Coast, Australia, where the researcher is employed. Bond University is a non-profit, private university (Table B university; Department of Education, 2019), which offers degree programmes to sub-bachelor, undergraduate, postgraduate and higher degree research students (Commonwealth Register of Institutions and Courses for Overseas Students, n.d.).

At the time of writing, there were approximately 3,500 full-time equivalent students enrolled at the university, including 39% international students (Times Higher Education, 2023). In terms of attrition, retention and success, data from the Department of Education (2023) indicates that the university experienced an attrition rate of 7.09% in 2020 amongst all bachelor level, undergraduate students. For sub-bachelor programmes, including diplomas, this rate increased to 14.75%. The retention rate in 2020 was 92.79% for bachelor degree students, and 84.75% for sub-bachelor students. Regarding success, the success rate in 2021 was 92.69% for bachelor degree students, yet dropped to 77.32% for sub-bachelor level students (Department of Education, 2023).

4.3 SAMPLE AND SAMPLING STRATEGY

Defined as “the source from which data are drawn to answer the research question(s)” (Perry, 2011, p. 56), the sample refers to the group examined empirically, whilst the target population is the group for whom the study concerns (Dörnyei, 2007; Riazi, 2016). Consequently, the following four samples representative of the population were identified as research participants: 1) students, 2) academic staff, 3) BELA users, and 4) external raters. The following sections describe the four participant groups, including background information, rationale for their inclusion in the study and sample size of each group.

4.3.1 STUDENT SAMPLE

The first group of participants, and the main sample investigated, comprised undergraduate university students enrolled in the subject, Core 1: Critical Thinking and Communication as part of a diploma or bachelor degree programme. These students are most affected by BELA, given they complete the task, with some required to undertake language development strategies. The cohort included students from both English-speaking and EAL/D backgrounds with the majority of participants in their first semester.

Student demographic and enrolment data, including degree programme and faculty of study, were obtained from a Senior Analyst in Bond University’s Office of Strategy, Planning and People, at the start of each semester during the data collection period. Academic performance data including BELA results, Major Essay marks and overall grade in the subject were collected via Blackboard. Students also completed a second BELA (i.e., BELA 2). In addition, a University Data Analyst (Learning & Teaching) provided students’ GPA after completing two semesters of university studies.

The student population comprised 1,246 students enrolled over five semesters. These semesters were as follows, with enrolment totals in parentheses: September 2018 semester ($n = 167$), January 2019 semester ($n = 415$), May 2019 semester ($n = 161$), September 2019 semester ($n = 148$), and January 2020 semester ($n = 355$). This population included students who were repeating the subject ($n = 70$). In order to accurately predict academic performance, the data of students repeating the subject was removed. Consequently, the final student research sample included 1,176 participants (94.3%). Student background data and academic performance data was available for each student. Sociodemographic characteristics of the sample population, including citizenship, language background, indigeneity, and disability, are presented in Table 4.1. The five most common degrees are included, as well as faculty/college at the University, comprising Faculty of Society and Design (FSD), Bond Business School (BBS), Bond University College (BUC), Health Sciences and Medicine (HSM), Law, and “Other” (e.g., Study Abroad).

Table 4.1

Sociodemographic characteristics of student sample

Characteristic	Sample	
	<i>n</i>	%
Gender		
Female	604	51.4
Male	572	48.6
Age		
18 years	347	29.6
17 years	231	19.7
19 years	184	15.7
20 years	98	8.3
Other ¹	314	26.7
Citizenship		
Australia	735	62.5
China	125	10.6
Canada	55	4.7
USA	25	2.1
India	23	1.9
Other ¹	213	18.1

Language background		
English	878	74.7
Mandarin	123	10.5
Arabic	27	2.3
Japanese	14	1.2
Portuguese	13	1.1
Other ²	121	10.3
Indigeneity status		
Non-Indigenous	1,144	97.3
Indigenous	32	2.7
Disability status		
No disabling cond.	1,150	97.8
Disabling cond.	26	2.2
Degree		
Law	99	8.4
Biomedical Sci.	88	7.5
Business (Bach.)	82	7.0
Construction Man.	66	5.6
Business (Dip.)	62	5.3
Other	779	66.2
Faculty/College		
FSD	329	28.0
BBS	280	23.8
BUC	240	20.4
HSM	154	13.1
Law	103	8.8
Other	70	5.9

Note. ¹ Other age included an age range of 21 to 47 years, and one student aged 16 years.

² Other citizenship included 47 others, such as Oman, Japan, Norway, and New Zealand.

³ Other language included 33 others, such as Norwegian, Russian, Cantonese, and Spanish.

⁴ Other degree programmes included 52 others, such as Bachelor of Psychological Science.

⁵ Other faculty included students for a semester and study abroad students.

Table 4.1 above illustrates the diverse sample of students included in the study. The age range of students was 16 to 47 years, with the mean age being 20.09 years ($SD = 4.64$ years). Most students were 18 years (29.6%), 17 years (19.7%) or 19 years (15.7%).

Regarding gender, slightly more females (51.4%) comprised the sample compared to males (48.6%). Most students held Australian citizenship (62.5%); however, it is important to note that Australian citizenship does not guarantee English is spoken at home. English was the most common language background (74.7%) followed by Mandarin, Arabic, Japanese and Portuguese. Notably, there were 33 other language backgrounds comprised in the sample, including Norwegian, Russian, Cantonese and Spanish, again demonstrating the diverse sample. A small percentage of students were Aboriginal and Torres Strait Islander students (2.7%), whilst a smaller percentage identified as being faced with a disabling condition (2.2%). Regarding degree programmes, a range of programmes can be seen, with Bachelor of Law (8.4%), Bachelor of Biomedical Science (7.5%) and Bachelor of Business (7%) the three most common degrees. Students from all faculties at the University were represented, with the Faculty of Society and Design (28%), Bond Business School (23.8%) and Bond University College (20.4%) having the highest representation.

In addition to student background and academic performance data, 110 students (9.4%) completed the student questionnaire. Concerning gender, more females ($n = 72$, 65.5%) completed the questionnaire compared to males ($n = 33$, 30.0%) with one student identifying their gender as “other”, and four students not responding to the question. Additional sociodemographic characteristics of the sample population completing the questionnaire, including age group, citizenship, and language background, are presented in Table 4.2 below.

Table 4.2

Sociodemographic characteristics of student questionnaire respondents

Characteristic	Sample	
	<i>n</i>	%
Gender		
Female	72	65.5
Male	30	30.0
Other	1	0.90
Age group		
16 – 18 years	41	39.4
19 – 21 years	34	32.7

22 – 24 years	14	13.5
25 – 28 years	7	6.4
29 – 31 years	3	2.7
32+ years	5	4.8
Citizenship		
Australia	75	70.8
Canada	7	6.6
Japan	7	6.6
China	4	3.8
Philippines	4	3.8
Other ¹	9	8.4
Language background		
English	88	83.8
Japanese	6	5.7
Mandarin	3	2.9
Other ²	13	11.8

Note. ¹ Other citizenship included six others, such as South Africa, Singapore, Sweden, and Thailand.

² Other language included eight others, such as Russian, Portuguese, Swedish, German, and Arabic.

Table 4.2 above illustrates the sociodemographic characteristics of the 110 student questionnaire respondents. Females are highly represented (65.5%). In terms of age, the majority of student respondents were in the 16 to 18 years age range (39.4%), followed by 19 to 21 years (32.7%). The citizenship comprised mostly Australian citizens (70.8%), followed by Canadians, Japanese, Chinese, Filipino and other nationalities (e.g., South African, Singaporean, Swedish and Thai). Finally, the majority of respondents used English at home (83.8%), followed by Japanese (5.7%) and Mandarin (2.9%) speakers. The sample of students completing the questionnaire is diverse and representative of the population.

4.3.2 OTHER SAMPLES: ACADEMIC STAFF, BELA USERS AND EXTERNAL RATERS

This section presents other samples included in the research. This comprises academic staff, who taught on the subject, Critical Thinking and Communication and also rated BELA essays. It also includes BELA users, key stakeholders at the university, and finally external raters.

The second group of participants were academic staff teaching on Core 1: Critical Thinking and Communication. It was necessary to gain feedback from these academics, as they are responsible for rating students' performance on BELA. Insight was, therefore, sought regarding areas including the BELA itself, follow up support mechanisms, the purpose and objectives of the assessment, as well as what factors, if any, affected their ratings. The academic staff sample comprised two lecturers, including the subject convenor, and seven tutors. In total, nine academic staff made up the academic staff sample. Of the nine academics, questionnaire data was obtained from eight academic staff (88.8%).

The third group of participants was other stakeholders who may use BELA results in their roles at the research site. This group was labelled, "BELA users". These stakeholders are considered users of BELA results, as the researcher sends them data regarding BELA completion and performance each semester. Specifically, on Monday, week 3 of the 14 week semester, BELA users, such as Associate Deans of Students and Academic Service Quality, are sent an Excel spreadsheet containing a list of students whose writing was identified as being Below satisfactory, as measured by BELA, as well as a list of students who did not complete BELA, possibly indicating lack of engagement. The BELA users sample was included, as it was important to obtain insight from these stakeholders concerning the consequences of the assessment. Similar to the academic staff sample, insight was gained regarding BELA, follow up support mechanisms and the purpose and objectives of the assessment. In total, 20 stakeholders were identified as the target BELA users population, as shown in Table 4.3 below. Questionnaire data was obtained from 19 BELA users (95%).

Table 4.3

Roles and number of staff identified as BELA users

Role at university	No. of staff
Deputy Vice Chancellors	2
Deputy Dean, Faculty of Law	1
Associate Deans of Student Affairs & Service Quality	4
Students and Academic Service Quality Managers	4
Director, Bond University College	1

Executive Manager, Bond University College	1
Student Transition and Retention Officer	1
Coordinator Academic Pathways Programs	1
Manager, Nyombil Indigenous Centre	1
Disability Support Officer	1
Core Curriculum Executive Officer	1
Business Academic Support Programme coaches	2
Total	20

As displayed in Table 4.3 above, the University staff identified as BELA users consisted of 20 members distributed across various roles. At the Executive level, there were two Deputy Vice Chancellors and 1 Deputy Dean for the Faculty of Law. At faculty level, the Associate Deans of Student Affairs & Service Quality and the Students and Academic Service Quality Managers both had four staff members each. Bond University College was represented by one Director and one Executive Manager. Amongst professional staff, there was one Student Transition and Retention Officer, one Coordinator for Academic Pathways Programs, one Manager of the Nyombil Indigenous Centre, one Disability Support Officer, and one Core Curriculum Executive Officer. Additionally, two coaches of the Business Academic Support Programme were identified as BELA users. BELA users receive data concerning BELA completion and performance each semester. They are considered key stakeholders; members of the University's Executive and Faculty Associate Deans can influence BELA policy, whilst professional staff can impact communication with students concerning language development.

The final group of participants were external raters. Two external language assessment experts, not associated with the University, participated in the study. The two raters, External rater 1 and External rater 2, were employed to ensure unbiased review of BELA and to minimise the possibility of a conflict of interest, given that other participants were associated with the research site. The external raters assessed approximately 10.2% of the BELA scripts ($n = 120$) and provided feedback on the measure and associated processes. This also enabled the researcher to evaluate BELA's inter-rater reliability. External raters'

participation was rewarded through financial remuneration from the researcher's higher degree research support funds. Details of the external raters, including recruitment, training procedures and indicative interview questions can be found in Appendix E. Upon completing online rater training, the researcher interviewed each external rater separately via Microsoft Teams to provide moderation feedback based on the six ratings and gain insight into the raters' initial observations regarding BELA (see Appendix E for indicative questions). A second semi-structured interview was conducted after rating had occurred via Microsoft Teams to clarify any discussion points that had arisen in the initial interview and to gain further, more detailed feedback about BELA and the associated processes. The interviews lasted approximately 30 minutes each and were recorded and transcribed using both Otter (audio) and Microsoft Teams (audio-visual).

In summary, the current study comprised the following four groups: 1) students, 2) academic staff, 3) BELA users, and 4) external raters. Students were the main sample investigated, as they are most affected by BELA, given they complete the task, with some required to undertake language development strategies.

The final student sample included 1,176 participants. The mean age of students was 20.09 years. There were slightly more females than males, with most holding Australian citizenship. Regarding language background, English was the most common followed by Mandarin, Arabic, Japanese and then Portuguese. Notably, 33 other language backgrounds comprised the sample. A small percentage of students were Aboriginal and Torres Strait Islander students, while a smaller percentage had indicated they were faced with a disabling condition. A range of degree programmes were studied by the sample with Bachelor of Law, Bachelor of Biomedical Science, and Bachelor of Business the three most commonly studied. Students from all faculties at the University were represented.

The second group of participants were two lecturers and seven tutors teaching on the subject, who rated students' performance on BELA. The majority of the academics responded to the questionnaire. Moreover, 20 stakeholders who may use BELA results in their different roles at the University were identified as BELA users. Importantly, BELA users represented a range of roles at the University including executive level, faculty deans and professional staff. Most BELA users responded to the questionnaire. Finally, two external raters participated in this research. These language assessment experts were employed to provide

objective review of BELA via semi-structured interviews and assess approximately 10% of students' BELA essays for inter-reliability analysis.

4.3.3 SAMPLING STRATEGY AND PROCEDURES

This study employed purposive sampling, defined as focusing on particular characteristics of participants to obtain a sample indicative of the wider, general population (Dörnyei & Taguchi, 2010; Woodrow, 2014). Specifically, convenience sampling, a purposeful sampling strategy (Perry, 2011), was used involving the selection of a conveniently available sample for the purposes of the research (Dörnyei, 2007; Riazi, 2016). In order to generalise findings, it is necessary to determine the sample is indicative of the target population (Perry, 2011). As such, identifying stakeholders who were affected by BELA (i.e., students and staff at the University) was purposive in that the researcher made the determination that the sample was the 'correct' one in terms of gathering data required to respond to the study's research questions.

4.3.4 RECRUITMENT OF PARTICIPANTS

This section presents the recruitment methods of each stakeholder group, beginning with the procedures of recruiting the student sample, followed by academic staff, BELA users and, finally, the external raters.

Regarding the recruitment of student participants, students were notified of the research each semester, over four semesters from September 2018 to, and including, January 2020. The researcher introduced the project and requested student participation during the week 12 lecture delivered by Academic Skills Centre Learning Advisors to assist students with their Major Essays for Core 1: Critical Thinking and Communication. During the workshop, the researcher invited students to participate in the research by: 1) completing the student questionnaire and/or 2) completing a second version of BELA, referred to as BELA 2. As a follow up strategy, all students enrolled in the subject, were again invited via email to participate in the research. A follow-up email was sent approximately two weeks later as a reminder. As enticing students to complete BELA 2 has proven difficult (Lydster & Brown, 2017); BELA 2 was incentivised by awarding 5 Beyond Bond points for completion. Beyond Bond is an activity-based, career development programme, a mandatory degree requirement, which encourages and recognises experiences students engage in, in addition to their studies. At the time of writing, students had to obtain 100 Beyond Bond points by the end of their

degree in order to graduate (Bond University, 2022a). No incentive was provided for completing the student questionnaire.

In terms of recruitment of academic staff participants, an email was sent to the identified academic staff population (i.e., nine individuals) in October, 2019. The communication explained the purpose and aims of the study, and provided a link to the relevant online questionnaire, with follow up emails sent individually two weeks later. This procedure resulted in the majority of staff participants, eight out of nine, completing the survey.

Recruitment of BELA users involved firstly sending an email to the entire BELA user population (i.e., 20 individuals) in August, 2019. Similar to the email sent to academic staff, the email contained an explanation of the purpose and aims of the study. A link to the relevant online questionnaire was included. A follow up email was also sent to BELA users individually in October, 2019. The majority of BELA users participated (19 BELA users).

Recruitment of external raters was accomplished via the researcher's network of language assessment professionals. The researcher's network comprised members of the Association for Language Testing in Australia and New Zealand (ALTAANZ) and the Association of Academic Language and Learning (AALL). The external raters were identified after considering the following desirable criteria: 1) having extensive experience in language assessment generally, 2) experience as an educator, particularly in the field of ALL or TESOL, 3) knowledge of applied linguistics, demonstrated by possessing a qualification in applied linguistics or TESOL, and 4) familiarity with PELAs. The identified candidates were contacted via email and two candidates accepted the offer to participate in April 2022.

4.4 ETHICAL CONSIDERATIONS

Approval to conduct the current study was gained from Bond University's Human Research Ethics Committee (Project CL03283) on 20 March, 2019. Informed consent, permission to conduct the study, assurances of anonymity and equality were all addressed and adhered to. All data used in the study was de-identified in accordance with the Australian National Data Service guidelines. De-identification involved the removal of any direct identifiers such as names or student identification numbers and replaced with "Student 1", "Tutor 1", "BELA user 1" and so forth, depending on the category of participant (i.e., student, academic or BELA user). All participants in the study will remain de-identified to all

researchers, employees of Bond University and any third parties who may be allowed access to the data.

4.5 RESEARCH TOOLS

Eight research tools for collecting data were employed. These included: 1) BELA, 2) Major Essay 3) overall grade in the subject, and 4) GPA after two semesters of university studies. In addition, the following three questionnaires were employed: 5) student questionnaire, 6) academic staff questionnaire, and 7) BELA users questionnaire. Finally, 8) Academic Skills Centre feedback survey, which included two items regarding BELA, was utilised in this study. The following sections detail the tools used.

4.5.1 BOND ENGLISH LANGUAGE ASSESSMENT (BELA)

The primary tool used in the current research was Bond English Language Assessment (BELA). BELA is a measure of academic essay writing skills, with the purpose to identify, early in their undergraduate degree programmes, students whose academic essay writing skills require support. BELA is considered an academic writing screening assessment, with potential to provide diagnostic information to individual students, academic staff and BELA users to inform learning and teaching. The test targets both EAL/D and English-speaking background students whose academic essay writing skills need improvement as early as possible in their degree studies.

BELA was created in 2014 by Learning Advisors at ASC in consultation with the convenor and lecturer of Core 1: Critical Thinking and Communication and the University's IT Learning Systems Project Manager. BELA was integrated as a low stakes assessment into the subject, in the third semester of 2014 as a compulsory homework task worth 2% towards the overall grade. The task was set in the first week of the semester, to be completed by the second week. The subject must be completed by all undergraduate students, with the exception of Bachelor of Medical Studies students. Therefore, BELA reaches most undergraduate students at the University. As found in a previous study, approximately 92% of students enrolled in the subject complete BELA (Lydster & Brown, 2017).

BELA requires students to consider a controversial statement about education, and then discuss it in the form of an academic essay. A word limit of between 300 and 400 words is given with a 60 minute-time limit. BELA has two versions, which differ only in the prompt provided. Version 1 requires students to argue the effectiveness of face to face versus online

education, whilst version 2 requires students to discuss whether exams are the best method of assessing students' knowledge (see Appendix G).

In terms of rating, BELA essays are assessed by a single rater for each tutorial class. The rating scale to assess students' writing skills was developed through consultation between Learning Advisors and academic staff and represents the valued skills and abilities concerning academic essay writing. The analytical scale encompasses "Organisation", "Linking and flow", and "Grammar and vocabulary" (see Appendix H) and is divided into two levels, Below satisfactory and Satisfactory. Possible scores obtained from applying the scale are 3, 5, 7 and 9. Students who are identified as having Below satisfactory academic essay writing skills, as measured by a score of 3 or 5 on BELA, are strongly recommended to attend a consultation with a Learning Advisor at ASC to discuss language development. Attendance was incentivised by providing students 2% towards their overall grade in the subject. Students scoring 7 or 9 are considered to have Satisfactory academic essay writing skills, however, may attend a feedback session.

Regarding mode of delivery, BELA is self-administered online via the subject's Blackboard site. Students can complete it anywhere, at any time, as it is not monitored by invigilators. The decision to administer BELA online considered the large number of students enrolled in the subject, on average, 280 students per semester, making it impractical to organise computer laboratories. Table 4.4 summarises the features of BELA.

Table 4.4*Features of Bond English Language Assessment (BELA)*

Feature	Key Questions	BELA
1. Target	Who is targeted?	All undergraduate students except Bachelor of Medical Studies students
2. Mandatory completion	Is the completion of the PELA compulsory for targeted students?	Compulsory homework task integrated into Core 1: Critical Thinking and Communication worth 2%. Students in Below satisfactory must attend ASC consultation to gain 2%.
3. Development	Where was the PELA developed and by whom?	Developed by Learning Advisors at ASC, Bond University's ALL unit. Staff are trained Cambridge (e.g., IELTS, FCE) and ISLPR examiners.
4. Target skills	What are the skills targeted by the PELA?	Academic essay writing
5. Mode of delivery	What is the mode of delivery of the PELA?	60 minute-online assessment. Can be accessed from anywhere and is not invigilated. A 300–400 word academic writing task (argument essay) with a 60-minute time limit.
6. Assessment	How is the PELA assessed and by whom?	Lecturers and tutors assess students' writing using three-category, analytic rating scale ("Organisation", "Linking and flow" and "Grammar and vocabulary") assessed on 2-point scale ("Below satisfactory" or "Satisfactory") Four possible scores. Scores of 3 and 5 are considered Below satisfactory; scores of 7 and 9 considered Satisfactory. Students in Below satisfactory must attend ASC consultation. No penalty for non-attendance.

Feature	Key Questions	BELA
7. On-campus support	What support mechanisms is the PELA linked to?	Feedback via rubric. ASC consultation(s) for detailed feedback. Language development plan including resources provided.

In summary, the purpose of BELA is to identify students whose academic writing skills are below satisfactory as early as possible in their undergraduate degree programmes, so that those identified are afforded additional support. All undergraduate students, except Bachelor of Medical Studies students are targeted. In line with other university PELAs, BELA is not a high-stakes gate-keeping test used for university admission. Integrated into the subject's curriculum and assessment, BELA is a compulsory homework task worth 2% towards the subject's overall grade. The PELA was developed by Learning Advisors at ASC, who are trained Cambridge (e.g., IELTS) and International Second Language Proficiency Ratings (ISLPR) examiners, in collaboration with the subject's convenor and a Learning Systems specialist. As a measure of academic essay writing, the task requires students to consider a controversial statement about education, and then discuss it in the form of an academic essay. As a response to academic integrity concerns, there are two versions of BELA. BELA is administered online, without the presence of invigilators. A single rater for each tutorial class rates students' BELAs. BELA raters are academic staff employed as tutors and/or lecturers on the subject; hence, raters are domain experts not language assessment experts. Students identified as having Below satisfactory academic essay writing skills, as measured by a score of 3 or 5 on BELA, are strongly recommended to attend a one-on-one consultation with a Learning Advisor at ASC to discuss academic language development. During this session, the Learning Advisor and the student review the essay and provide detailed feedback in terms of their strengths and weaknesses in academic essay writing. Learning resources are recommended, and a personalised plan of academic language and literacies development is discussed.

4.5.2 OTHER TOOLS: BELA 2, MAJOR ESSAY, OVERALL GRADE, & GPA

This section presents other tools utilised in the research. Other tools include BELA 2, the second version of BELA. It also comprises the Major Essay for the subject, Critical Thinking and Communication, as well as the overall grade in the subject and grade point average or GPA. The three questionnaires and Academic Skills Centre survey were also used in the research.

The second research tool used in this study comprised BELA 2. At the conclusion of each semester, students were invited to complete a second BELA. In total, 68 students completed BELA 2. Students who had completed version 1 of BELA at the beginning of the semester completed version 2 at the end of the semester and vice versa. Students completed

BELA 2 via the ASC Blackboard site, as BELA had no bearing on students' grades in the subject. Academic staff were not involved in the rating of BELA 2 due to time and budget constraints. For consistency, the researcher also rated the students' initial BELAs to determine whether their academic essay writing had developed over the course of one semester of study.

The third research tool was the Major Essay. Students enrolled in Core 1: Critical Thinking and Communication, complete one major academic essay as an assignment. The Major Essay, worth 25% of the overall subject grade, has a word limit of 1,200 words (+/- 10%) and is due at the end of the semester. Students must present an argument for believing or disbelieving in a certain topic or field, for instance, chiropractic, ear candling or reflexology. Students are randomly given a topic in week 8 and they are required to write an academic essay, incorporating literature and evidence to support their arguments. The rubric for the Major Essay comprises the following criteria: Introduction, Arguments, Organisation, Research and referencing, Academic style, editing and proofreading, Critical thinking consideration and Argument (see Appendix I). Major Essays are assessed by academic staff on a 4-point scale from Very Good/Excellent to Does not meet the minimum requirements.

The fourth tool was students' overall grades in Core 1: Critical Thinking and Communication, utilised as a measure of academic performance. These data were obtained via the subject's learning management system, Blackboard. Overall grades are scores presented as percentages out of 100.

GPA is defined as "the overall averaged subject-matter grade point" (Strickland, 1998, p. 3). At the research site, GPA is defined as the average of the grades that are achieved by students in all of their subjects, weighted by the value of credit attached to each subject (Bond University Student Administration, n.d.). At Bond University, GPA is measured on a continuum from 0 to 4. A GPA between 0 and .99 indicates the student is failing, whilst a GPA above 1 indicates the student is averaging at least a pass (1.0 = Pass, 2.0 = Credit, 3.0 = Distinction, 4.0 = High Distinction). For the purpose of this study, students' cumulative GPA data after two semesters of study were used as a measure of academic performance.

4.5.3 QUESTIONNAIRES

Questionnaires were chosen to elicit attitudes and perceptions amongst students, academic staff and BELA users. Questionnaires are efficient, versatile and reliable (Creswell,

2008; Dörnyei, 2007; Dörnyei & Taguchi, 2010) and have proven to be effective in ascertaining stakeholders' attitudes about language learning (Bernat & Gvozdenko, 2005; Tanaka & Ellis, 2003). For the current study, a standardised measure of stakeholders' attitudes and perceptions towards a PELA was non-existent; therefore, three questionnaires were developed by the researcher, based on the extant literature on validating English language assessments (e.g., Elder & Erlam, 2001) and the validity framework guiding this study (Knoch & Elder, 2013).

The first questionnaire was a 49-item questionnaire designed to be used with student participants to gain insight into their experiences with BELA and associated support available to them (see Appendix J). The second was a 45-item questionnaire used with the subject's academic staff in order to review the whole BELA process, including the assessment itself and connected support mechanisms (see Appendix K). The third was a 35-item questionnaire designed for BELA users, including Deputy Vice Chancellors and Associate Deans (see Appendix L). Each questionnaire was web-based using Qualtrics, due to the expediency and efficiency in reaching the target population and collecting data in a quick and cost-effective manner (Creswell & Guetterman, 2019; Rea & Parker, 2014).

In the three questionnaires, the first page contained the explanatory statement, and if participants agreed to proceed, they were advised to tick a box indicating their consent to participate. The three questionnaires contained a series of items asking participants to indicate their level of agreement using a 4-point Likert scale (1 = *Strongly disagree*, 4 = *Strongly agree*). An even numbered scale was chosen, without a neutral point, to avoid participants "sitting on the fence", as it has been reported that participants who are less motivated to complete a questionnaire are more inclined to respond by selecting the mid-point or neutral point on a scale (Krosnick, 1999; Krosnick et al., 2005). Each item was characteristic (i.e., items express either a positive or negative attitude regarding the topic of investigation), with neutral and extreme items not included. The majority of items in each questionnaire were closed ended questions, as they best facilitate comparisons amongst participants, maximise clarity and limit the number of extraneous responses (Rea & Parker, 2014). However, open ended questions to gain qualitative feedback from stakeholders were included. The time required to complete the questionnaire was aimed to be less than 20 minutes in accordance with Dörnyei and Taguchi's (2010) recommendations.

To identify risks to measurement error, the student questionnaire was piloted with two cohorts of students enrolled in the subject over two semesters (January 2019 and May 2019). The questionnaire was introduced to students in week 12 of the two semesters during an integrated workshop. Students were invited to participate via accessing a hyperlink or QR code. In total, 40 students completed the pilot questionnaires (17 in January semester and 23 in May semester). During the pilot, student respondents on average took 5 minutes and 34 seconds to complete the questionnaire, well within the recommended maximum duration of 30 minutes (Dörnyei, 2007). The student questionnaire was the only questionnaire piloted for two reasons. Firstly, the majority of participants were predicted to comprise the student participant group, given the population was much larger. Secondly, many of the items in the student questionnaire were similar to items in the other two questionnaires; therefore, it was considered redundant to pilot all three questionnaires. The student questionnaire was again presented to students in the September 2019 and January 2020 semesters, with an additional 70 students responding to the student questionnaire. No incentive was given. In total, 110 students, eight academic staff and 19 BELA users completed the respective questionnaires. Table 4.5 below displays the reliability of each questionnaire, as per the five inferences of the validity framework. Overall, the questionnaires displayed medium to good reliability; however, some inferences presented poor reliability, which will be discussed as limitations.

Table 4.5

Questionnaire reliability

Inference	Student questionnaire	Academic staff questionnaire	BELA users questionnaire
Evaluation	.60	.71	.65
Generalisability	.18	.76	NA
Extrapolation & Explanation	.66	.68	.39
Decisions	.56	-.75	.68
Consequences	.76	.39	.76

Note. The BELA users questionnaire only contained one item regarding Generalisability.

4.5.4 ACADEMIC SKILLS CENTRE FEEDBACK SURVEY

Bond University's academic language and learning unit, ASC seeks feedback from students who have attended a consultation with a Learning Advisor. The feedback survey is a Survey Monkey survey comprising nine questions. Feedback is sought on topics, such as whether they were able to book an appropriate time at the centre (e.g., "Generally, I was able to book a time I wanted"), if the support was beneficial (e.g., "Overall, I found the consultation with Academic Skills Centre useful"), as well as specific questions related to the individual Learning Advisor the student had consulted with (e.g., "His/her knowledge of writing, grammar & academic skills was good"). In the context of BELA and for the purposes of the current investigation, to gain further insight into students' perceptions of assessment and the associated language support offered, two items are included in the ASC feedback survey. The first item asks students whether they had completed BELA (i.e., "Did you do the BELA writing task for Core 1 this semester?"). Students who respond, "yes" to this question, were then prompted to give feedback about their experience with BELA using a comment box response question (i.e., "Tell us more about your experience?"). This survey is sent yearly to all students who had attended at least one consultation at ASC. This was achieved by downloading a spreadsheet of all student appointments in the relevant semester and emailing students who had attended a consultation at ASC, inviting them to complete the survey by clicking on the link in the email. Between 2017 and 2022, 314 students, who attended at least one ASC appointment, responded to the survey. Regarding BELA, 84 students responded that they had completed the measure, 217 had not and 13 skipped the question. Of the 84 student respondents who had completed BELA, 52 students responded to the open-ended item requesting feedback on their experience with BELA.

4.6 DATA ANALYSIS PROCEDURES

This section details the analysis procedures undertaken. It begins by introducing the data analysis procedures for the quantitative data, including student background data, academic performance data, questionnaire data (i.e., student, academic staff and BELA users questionnaires) and BELA reliability data. The section then documents the analysis procedures for the qualitative data utilised, including the questionnaires, ASC feedback surveys and semi-structured interviews with external raters.

4.6.1 QUANTITATIVE DATA ANALYSIS

This section presents the data analysis procedures for the quantitative data. These data included student background data, academic performance data, questionnaire data, and BELA reliability data.

First, student background data was analysed by transferring coded data from Excel into IBM SPSS version 28 by the researcher. Data was checked for errors and analyses were run. Specifically, using language background and BELA results (i.e., Below satisfactory and Satisfactory), chi-square tests were run to determine if BELA unfairly favoured certain groups of students. To determine what factors, if any, could predict academic performance, binary logistic regression was conducted.

Second, academic performance data was analysed by firstly transferring coded data from Excel into IBM SPSS version 28 by the researcher. Data was checked for errors and analyses were run. Specifically, Point Biserial correlation was run between BELA categories (Below satisfactory and Satisfactory) and academic performance in Major Essay, overall grade in the subject and GPA after two semesters. Point Biserial correlation was also run to determine if students who were categorised as having Below satisfactory academic essay writing and did not attend ASC consultations performed worse academically (i.e., Major Essay, overall grade and 2nd semester GPA) than students who were in the same category but did attend such consultations. Regarding BELA 2, comparative analysis was conducted via Excel. The scores of students regarding BELA 2 for each band (i.e., 3, 5, 7, and 9) were analysed to gauge the proportion of students who improved by at least one band in BELA 2. Additionally, comparison was made between students who had attended consultations at ASC compared to those who had not. To determine what factors, if any, could predict academic performance, binary logistic regression was conducted.

Third, questionnaire data was exported from Qualtrics to IBM SPSS version 28. Data was checked for errors and analyses were run. Specifically, independent samples t-tests were run to determine if significant differences of opinion were apparent between stakeholders regarding common items (e.g., “The criteria [1. organisation, 2. linking and flow, and 3. grammar and vocabulary] capture aspects of performance most relevant to academic essay writing”). Additionally, one-way ANOVAs were used to compare stakeholder responses for items, such as “The time limit (60 minutes) was enough time to complete BELA.”

Finally, in order to determine the reliability of BELA, it was necessary for the researcher to assess a sample of BELA scripts/essays and compare the ratings with the ratings of academics on the subject. In total, 180 BELAs (16.13% of the population) written by 180 unique students were assessed by the researcher and compared to the initial ratings of the nine academic staff who taught on the subject. Specifically, 20 BELAs were reassessed per academic, including 10 version 1 BELAs and 10 version 2 BELAs for tutors 1 to 8. For tutor 9, 20 version 1 BELAs were reassessed as tutor 9 did not teach in a semester in which version 2 was employed. The chosen scripts were selected via stratified random sampling. The strata of concern were the nine academic staff that had rated students' BELA essays. The total number of scripts rated by each academic for both versions of BELA were divided by 10 and then scripts were selected accordingly. For example, if an academic had rated 70 version 1 scripts, the first BELA and every 7th BELA thereafter was chosen. This approach was taken to ensure an adequate distribution of BELA essays to re-assess and be representative of the population (Riazi, 2016). Two additional raters, External rater 1 and External rater 2, were employed as participants to assess approximately 10% of the BELA essays ($n = 120$) to ensure unbiased review of BELA. External rater 1 assessed BELA version 1 scripts, whilst external rater 2 assessed BELA version 2 scripts. In total, external rater 1 assessed 52 version 1 and external rater 2 assessed 68 version 2 BELAs. Selection of the 120 BELA scripts involved using all of the scripts for which the original raters (i.e., academic staff) and the second rater (i.e., the researcher) were not in complete agreement ($n = 94$). The remaining 26 BELA scripts were selected using simple random sampling to select every third BELA of the remaining 86 scripts that had been rated by the original raters and the researcher.

Analyses of the BELA reliability data collection involved ratings being downloaded into an Excel spreadsheet, then combining the ratings of second, third and external raters into one file. The Excel spreadsheet was then converted into an SPSS file for analysis. Data was cross checked in Excel and the SPSS file. Analysis of the inter-rater reliability of BELA was then performed. This involved calculating Cohen's kappa to determine level of agreement amongst the raters' scores. To determine which criterion included higher/lower level of agreement, Cohen's kappa was also used. Subsequently, ratings were dummy coded to determine levels of agreement regarding the categorisation of students' writing (i.e., Below satisfactory or Satisfactory), as these categories were of key concern.

4.6.2 QUALITATIVE DATA ANALYSIS

This section presents the data analysis procedures for the qualitative data. These data were obtained via open-ended items in the three Qualtrics questionnaires (i.e., student questionnaire, academic staff questionnaire, and BELA users questionnaire), the ASC feedback survey, and semi-structured interviews with external raters. Responses to the open-ended items from the three questionnaires were downloaded from Qualtrics and then uploaded into NVivo 12 Pro for analysis. Regarding the ASC feedback survey, comments were first downloaded from the survey platform, Survey Monkey, as an Excel spreadsheet and then uploaded to NVivo 12 Pro. Finally, analysis of the semi-structured interview data involved firstly uploading the transcripts of the interview recordings as Word documents into NVivo version 12 Pro. External raters were deidentified and were referred to as “External rater 1” and “External rater 2” respectively. For each data type, the researcher coded the data line by line in NVivo 12 Pro to determine common themes (e.g., “implementing scoring”, “difficulty using rubric”).

4.7 RESPONDING TO THE RESEARCH QUESTIONS

This section synthesises the data sources and statistical analyses applied to respond to the study’s research questions. The first research question was: Can a satisfactory validity argument be presented for the Bond English Language Assessment (BELA)? This research question comprised five sub-questions. The second research question was a stand-alone question: What factors predict academic performance of university students? The following sections are divided based on the inferences of Knoch and Elder’s (2013) PELA validity framework for RQ1a to RQ1b, and finally, the second research question (RQ2).

4.7.1 EVALUATION INFERENCE

Research question 1a concerned how adequately BELA scores reflect behaviours of test takers. The purpose of this question was to determine whether the BELA rubric provides evidence of academic essay writing skills, whether test administration conditions are acceptable, and whether statistical characteristics of BELA, including different test forms, are appropriate to make decisions (Chapelle, 2008; Knoch & Elder, 2013). This question is aligned with the Evaluation inference of the framework comprising five warrants. Table 4.6 summarises the type of evidence sought to respond to RQ1a.

Table 4.6*Warrants and evidence for the Evaluation inference for BELA*

Evaluation inference	
RQ1a: Are BELA scores adequate reflections of observed behaviours?	
Warrants	Sources of supporting evidence
1. Scoring criteria and rubrics capture relevant aspects of performance.	External rater feedback Academic staff questionnaire responses
2. Raters can implement scoring procedures consistently.	Inter-rater reliability analysis External rater feedback Academic staff questionnaire responses Student questionnaire responses
3. Test administration conditions are clearly articulated and appropriate.	Student questionnaire responses Academic staff questionnaire responses BELA users questionnaire responses
4. Instructions and tasks are clear to all test takers.	Student questionnaire responses Academic staff questionnaire responses
5. Test is pitched at appropriate difficulty level and test tasks/items discriminate consistently between more and less able candidates.	Test difficulty and discrimination analyses External rater feedback

As displayed in the table, Warrant 1, which requires the scoring rubric to capture relevant aspects of performance, involved evaluation by academics via the academic questionnaire (i.e., “The criteria [1. organisation, 2. linking and flow, and 3. grammar and vocabulary] capture aspects of performance most relevant to academic essay writing.”). It also involved the external raters reviewing the scoring rubric via interview. Warrant 2, requiring raters to be able to implement consistent scoring procedures, involved analysing the item in the academic staff questionnaire which asked respondents how confident they were in applying the scoring procedures (i.e., “I can consistently apply the BELA criteria”). Students were also asked whether they believed their writing was correctly rated via the student questionnaire (i.e., “My writing was correctly rated”). Qualitative insight from external raters into their ability to apply the rubric was also obtained. Finally, inter-rater reliability analysis, using Cohen’s kappa coefficient with the ratings of the researcher (i.e., the second rater; $n =$

180), as well as two external raters (External rater 1, $n = 52$; External rater 2, $n = 68$), compared to the original ratings, was used to determine support for this warrant.

Furthermore, in regards to Warrant 3, necessitating test administration conditions be clear and appropriate, students’ reflection on the process they undertook in completing the test, as well as review of administration and procedures (e.g., quiet environment, sufficient time limit, technical difficulties), was ascertained via the student questionnaire, academic staff questionnaire, and BELA users questionnaire (i.e., “I had no technical difficulties when completing BELA”). Here, independent samples t-tests were used to compare responses of stakeholders. Responding to warrant 4, requiring clarity of instructions and tasks, involved analysis of student and academic staff responses to the relevant item in the student questionnaire (i.e., “The instructions for BELA were clear”). Finally, Warrant 5, which states the test is appropriately pitched in terms of difficulty and can discriminate consistently based on students’ abilities, required statistical analysis. Assessing test difficulty and discrimination power involved dummy coding BELA scores as Below satisfactory (scores of 3 and 5 = 0) and Satisfactory (scores of 7 and 9 = 1). Test difficulty and discrimination were then determined separately for each BELA version in accordance with Haladyna (1999) and Kelley (1939). It also involved gaining feedback from external raters.

4.7.2 GENERALISABILITY INFERENCE

Research question 1b focused on the generalisability of BELAs two forms (version 1 and version 2). The purpose was to assess whether BELA results are consistent across contexts (Knoch & Elder, 2013). This question aligns with the Generalisability inference, made up of four warrants. Table 4.7 details the warrants and evidence gathered to respond to RQ1b.

Table 4.7

Warrants and evidence for the Generalisability inference for BELA

Generalisability inference	
RQ1b: Does BELA yield results consistent across assessment contexts?	
Warrants	Sources of supporting evidence
1. Different test forms are parallel in design.	Researcher’s review of test forms External rater feedback

Generalisability inference

RQ1b: Does BELA yield results consistent across assessment contexts?

2. Appropriate equating procedures are used to ensure equivalent difficulty across test forms.	Pearson's chi-square test of contingencies
3. Sufficient tasks are included to provide stable estimates of test taker ability.	Comparison of psychometric properties of the test as compared to trial population External rater feedback
4. Test administration conditions are consistent.	Academic staff questionnaire Student questionnaire responses

Table 4.7 above illustrates that Warrant 1, necessitating different versions of BELA be parallel in design, required comparison of the two forms of BELA by the researcher and external raters. Warrant 2, ensuring that different test forms entail equivalent difficulty, involved the use of Pearson's chi-square test of contingencies to determine if there was a difference in test results (Below satisfactory; Satisfactory) between the two versions of BELA. It also involved analysing feedback from the external raters. Warrant 3, which focuses on sufficient tasks being included, again involved the use of Pearson's chi-square test of contingencies to determine if there was a difference in test results between a pilot sample and the current sample. Finally, Warrant 4, requiring test administration conditions remaining consistent, involved review of procedures via both the student and academic questionnaires.

4.7.3 EXPLANATION AND EXTRAPOLATION INFERENCE

Research question 1c comprised two questions. It first of all sought to determine whether BELA provides information on students' skills, knowledge and characteristics keeping within the understanding of academic ELP. Secondly, it focused on whether BELA is an adequate proxy for tasks that are performed in domain of academia. Overall, this question concerned how BELA tasks and the elicited language compare to task types and language used in the academic domain. It also concerned the accuracy to which BELA scores predict language performance in academia or in other assessments that have been designed to assess academic English (Knoch & Elder, 2013). RQ1c aligns with the Explanation and Extrapolation, made up of six warrants. Table 4.8 summarises the warrants and evidence sought for RQ1c.

Table 4.8*Warrants and evidence for the Explanation and Extrapolation inference for BELA*

Explanation and Extrapolation inference	
RQ1c: Does BELA provide information on test takers' skills, knowledge and characteristics that keep within the understanding of academic ELP? Is BELA an adequate proxy for tasks performed in the academic domain?	
Warrants	Sources of supporting evidence
1. Test results are good predictors of language performance in academic domain.	Point Biserial correlation analysis using BELA scores, scores on the Major Essay, overall grades in subject and GPA after two semesters of study
2. Characteristics of test tasks are similar to those required of students in the academic domain (and those in the language development courses students are placed in).	Student questionnaire responses Academic staff questionnaire responses Assessment data at research site External rater feedback
3. Linguistic knowledge, processes, and strategies employed by test takers are in line with theoretically informed expectations and observations of what is required in the corresponding academic context.	Student questionnaire responses Academic staff questionnaire responses
4. Scores derived from the test provide sufficient information about candidates' academic language proficiency (i.e., no construct under-representation).	Academic staff questionnaire responses Student questionnaire responses External rater feedback
5. Performance on PELA relates to performance on other assessments of academic language proficiency.	Data not obtained
6. Tasks do not unfairly favour certain groups of test takers.	Student questionnaire responses Academic questionnaire responses External rater feedback

Explanation and Extrapolation inference

RQ1c: Does BELA provide information on test takers' skills, knowledge and characteristics that keep within the understanding of academic ELP? Is BELA an adequate proxy for tasks performed in the academic domain?

Statistical analysis of scores based on language background

To summarise Table 4.8, Warrant 1, determining whether BELA results are adequate predictors of language performance at university, involved correlational analysis between BELA scores and performance in language related academic tasks, including the Major Essay in the subject, Core 1: Critical Thinking and Communication. This analysis involved determining series of Point Biserial correlation. Warrant 2, necessitating similarity between the BELA task and tasks students are required to perform at university, required students, academic staff and BELA users' review of BELA in terms of whether BELA captures the relevant aspects of tertiary level writing via the three questionnaires (i.e., "BELA is similar to other university assessment tasks). This also involved analysis of qualitative feedback from the external raters. Furthermore, the common assessment types at the research site were analysed. Warrant 3, which focuses on consistency between students' linguistic knowledge, processes and strategies utilised when completing BELA and what is required within academia, included analysis of responses to questionnaire items that focused on students' strategies employed when completing BELA, as well as review by the external raters.

Furthermore, Warrant 4, which involves determination of whether BELA scores provide sufficient information about students' academic language proficiency, included academic staff review of BELA obtained via questionnaires. Insight was also obtained from students as to whether they believed their writing in BELA was representative of their academic writing abilities (i.e., "My writing in BELA was a good example of my academic writing ability"). Finally, external raters' review via semi-structured interviews was analysed. Warrant 5, which states that students' performance on BELA is related to their performance on other assessments of academic language proficiency, was not investigated due to a lack of available data. No data was available on students' scores in other measures of language proficiency, such as IELTS. Had scores on other language assessments been available, such data would have been of limited value as the majority of students in the target population entered university without the necessity of obtaining language proficiency score. Finally,

warrant 6, which required that BELA does not unfairly favour certain groups of students, involved analysing feedback from students as test takers via the student questionnaire (i.e., “Some students have an unfair advantage over other students when completing the BELA task”), as well as insight from academic staff through the academic staff questionnaire (i.e., “BELA is a fair task.”), and external raters through semi-structured interviews to determine whether they viewed BELA as a fair task. Additionally, the BELA performance dependant on language background (i.e., English-speaking background in comparison to EAL/D students) was analysed.

4.7.4 DECISIONS INFERENCE

Research question 1d deviated from BELA itself as a measure (interpretation of scores), to results and recommendations after the assessment was completed and scores communicated (utilisation of scores). Here, the concern was how appropriate and how well communicated decisions based on BELA scores were (Knoch & Elder, 2013). Thus, evidence was required to demonstrate that BELA results are appropriate for providing advice to students concerning whether they may need to develop their academic essay writing skills and how (Read, 2015a). RQ1d aligns with the Decision inference and contains six warrants. Table 4.9 documents the warrants and evidence sought to respond to RQ1d.

Table 4.9

Warrants and evidence for the Decisions inference for BELA

Decisions inference	
RQ1d: Are decisions based on BELA scores appropriate and well communicated?	
Warrants	Sources of supporting evidence
1. Students are correctly categorised based on their BELA scores.	Analysis of Major Essay results and GPA based on BELA scores
2. The test results include feedback on test performance and a recommendation.	Student questionnaire responses Academic staff questionnaire responses BELA users questionnaire responses ASC feedback survey

Decisions inference

RQ1d: Are decisions based on BELA scores appropriate and well communicated?

3. The recommendation is closely linked to on-campus support.	Student questionnaire responses BELA users questionnaire responses
4. Assessment results are distributed in a timely manner.	Student questionnaire responses BELA users questionnaire responses
5. The test results are available to all relevant stakeholders.	BELA users questionnaire responses
6. Test users understand the meaning and intended use of the scores.	Student questionnaire responses Academic staff questionnaire responses BELA users questionnaire responses

In summary, Warrant 1, which requires correct categorisation of students based on BELA scores, involved analysing academic outcomes of students. One-way ANOVA was used to determine whether a significant difference existed between students scoring low and high on BELA for the Major Essay and GPA after two semesters. Warrant 2, necessitating feedback on BELA performance and a recommendation based on results, required feedback from students and academics through the relevant questionnaires/survey. Specifically, items were included that investigated whether students received a recommendation (“I received a recommendation/recommendations based on my BELA”), the amount of feedback provided (i.e., “Students are given sufficient feedback on their academic writing as part of the BELA process”), and accessibility of feedback (i.e., “It was easy to access feedback on my BELA”). Warrant 3, which requires recommendations to be linked to on-campus support, involved analysing feedback from the student questionnaire (i.e., “Recommendations provided to students according to their BELA scores are linked to the University’s on-campus support” and “Academic writing support is available at the University”).

In addition, Warrant 4, requiring results to be delivered in timely manner, also involved analysing responses to the student questionnaire (i.e., “I received the results on my BELA quickly”) and BELA users questionnaire (i.e., “BELA results are communicated to me in a timely manner”). Warrant 5, requiring BELA results being available to all stakeholders, involved analysis of feedback from BELA users via questionnaire (i.e., “All university

stakeholders who can influence students’ retention are provided with BELA data.”). The final warrant, warrant 6, concerning all BELA stakeholders understanding the meaning and intended uses of BELA scores, involved analysing the feedback from all BELA stakeholders via the three questionnaires. Specifically, this involved analysing items regarding the lower scores’ meaning (i.e., “A score of 3 or 5 on BELA indicates that a student’s academic writing is below satisfactory”) and then higher scores’ meaning (i.e., “A score of 7 or 9 indicates that a student’s academic writing is satisfactory”). Additional items were included in the student questionnaire to determine if they understood the meaning of their BELA results (i.e., “If I didn’t receive an email from my lecturer or tutor, it meant my writing was satisfactory” and “I do not know if my writing was satisfactory or unsatisfactory”).

4.7.5 CONSEQUENCES INFERENCE

Finally, research question 1e focused on the consequences of utilising BELA and whether the decisions based on BELA results are of benefit to stakeholders. Like the Decisions inference (RQ1d), the Consequences inference is governed by university policy and the overall context in which BELA is integrated (Knoch & Elder, 2013). Here, evidence of beneficial outcomes for students as test-takers, as well as Bond University overall, was required (Read, 2015a). This question aligns with the Consequences inference, which includes eight warrants. Table 4.10 summarises the warrants and evidence sought to respond to RQ1e.

Table 4.10

Warrants and evidence for the Consequences inference for BELA

Consequences inference	
RQ1e: Are the consequences of using BELA and the decisions informed by BELA beneficial to all stakeholders?	
Warrants	Sources of supporting evidence
1. All target test takers sit the test.	Internal data from Core 1 subject site (i.e., completion rates)
2. The test does not result in any stigma or disadvantage for students.	Student questionnaire responses Academic questionnaire responses BELA user questionnaire responses ASC feedback survey responses External rater feedback

Consequences inference

RQ1e: Are the consequences of using BELA and the decisions informed by BELA beneficial to all stakeholders?

	ASC feedback survey
3. Test takers' perceptions of the test and its usefulness are positive.	Student questionnaire responses ASC survey responses BELA user questionnaire responses
4. The feedback from the test is useful and directly informs future learning.	Student questionnaire responses Academic questionnaire responses BELA user questionnaire responses ASC feedback survey responses External rater feedback
5. Students act on the test recommendation.	Internal data from ASC (i.e., uptake rates)
6. Follow-up language development options provided for students is appropriate.	Student questionnaire responses Academic questionnaire responses BELA user questionnaire responses ASC feedback survey responses
7. Learners taking up support options improve their English over the course of their studies.	Matched sample analysis to determine gains on BELA 2
8. Students who fail to act on test recommendations are more likely to struggle in their academic studies.	Correlational analysis using internal data from ASC (i.e., uptake rates) and academic performance data

In sum, Warrant 1, which requires that all identified students complete BELA, involved analysis of completion data from internal records obtained via the Core 1: Critical Thinking and Communication learning management site (i.e., Blackboard). It also involved analysing reasons for non-completion through the student questionnaire. Warrant 2, concerned with the avoidance of stigma and/or disadvantage for all students, required analysis of multiple data types, including the student questionnaire (i.e., "I felt disadvantaged when completing BELA"), academic staff questionnaire (i.e., "There is a stigma attached to performing poorly on BELA") and BELA user questionnaire responses. It also necessitated analysis of the ASC feedback surveys and external rater feedback via semi-structured

interviews. Utilising these data aimed to elicit insight concerning perceptions of whether BELA stigmatises or disadvantages certain individuals or groups of students. Warrant 3, which necessitates students perceiving BELA and its usefulness in a positive manner, was evaluated by analysing student feedback from the student questionnaire (i.e., “Completing BELA has been beneficial for me” and “BELA benefits other students”), academic staff questionnaire, and BELA users questionnaire (i.e., “Completing BELA is beneficial for students”) and the ASC feedback surveys. Warrant 4, focusing on the usefulness of feedback from BELA and that it informs future learning, involved analysing student questionnaire responses (i.e., “The feedback I received after completing my BELA has helped me to develop my academic writing”) and the ASC feedback surveys. Feedback from academics and BELA users was also ascertained from the relevant questionnaires (i.e., “Students’ writing in BELA can inform instruction”). The external raters also provided insight via semi-structured interviews.

Moreover, Warrant 5, requiring students act on recommendations as a consequence of their BELA result, involved analysing uptake rates through internal data of students utilising the ALL unit, ASC. Warrant 6, concerning the appropriacy of follow-up academic language development options, required analysing student feedback from the student questionnaire (i.e., “As a consequence of attending Student Learning Support, my academic writing has improved” and “My English has improved as a result of attending Student Learning Support”), academic questionnaire (i.e., “Students English improves as a result of attending Student Learning Support”) and BELA users questionnaire (i.e., “As a consequence of attending Student Learning Support, students’ academic writing improves”), as well as the ASC feedback surveys. Items were also included in the student and academic questionnaire to determine whether these stakeholders believed that feedback did not assist in developing academic writing (i.e., “Even though I received feedback and support, my academic writing has not improved”). Warrant 7, concerning the development of academic essay writing skills amongst students who utilised language development options, required investigating the relationship between student performance on BELA at the start of the semester and the optional BELA 2 at the conclusion of the semester to determine any writing gains. Finally, Warrant 8, necessitating an analysis of the outcomes of students who do not act on recommendations regarding language development, required examining internal data from ASC and then analysis of students in the Below satisfactory BELA category who did not attend learning consultations and their performance on: 1) Major Essay, 2) overall grade in

the subject, and 3) GPA after two semesters. This was then compared to the performances of students who had also been categorised as having Below satisfactory academic essay writing but who had attended ASC using t-tests.

4.7.6 WHAT FACTORS PREDICT ACADEMIC PERFORMANCE OF UNIVERSITY STUDENTS?

The second research question investigated whether BELA scores and/or additional factors (i.e., demographic factors and academic performance data) were able to predict the academic performance of undergraduate students. This research question was posed in response to the overarching aim of the thesis, to identify what factors, if any, could predict academic performance at university. Here, academic performance was based on grade outcomes. Consequently, the academic performance of students was categorised into two groups: 1) students failing (i.e., scoring below 1.00 GPA after two semesters) and 2) students achieving high academic achievement (scoring 3.00 GPA and above after two semesters).

In order to determine whether BELA or other student demographic and/or academic performance data could predict students failing, binary logistic regression analysis was employed. Specifically, to estimate the probability of a student failing after two semesters as indicated via GPA of below 1.0 (e.g., .99), which indicated a failing average, a logistic regression was conducted. The probability was estimated depending on student demographics (i.e., gender; age; nationality; language background; faculty; country of origin; indigeneity; disability), as well as academic performance data, which were both added to the model (i.e., overall grade in Core 1; uptake of support; BELA result). Additionally, to determine if BELA or other student demographic and/or academic data could predict high academic achievement at university, defined as achieving a Distinction or above, a second logistic regression analysis was run. The logistic analysis to determine high achievement at university again followed the aforementioned steps, however, GPA 3.0 and above was used to indicate high achievement of Distinction average or above.

4.8 SUMMARY OF METHODOLOGY

The current study employed a mixed methods approach focusing on quantitative data and including complementary qualitative data. The quantitative data comprised: 1) student background data, 2) academic performance data 3) questionnaire data, and 4) BELA reliability data. Three categories of qualitative data were obtained: 1) questionnaire data, 2) ASC feedback survey, and 3) semi-structured interviews with two external raters.

Four groups were included in the sample. The main sample was undergraduate students enrolled in the subject, Core 1: Critical Thinking and Communication. The student sample was considered the main sample, as students take BELA and are most affected by it. The student sample comprised 1,246 students who were enrolled over five semesters. After removing the data of students who were repeating the subject, the student sample included 1,176 individual students. The second group of participants was academic staff teaching on the subject. Their feedback regarding areas such as the BELA itself, the objectives of the assessment, and what factors, if any, affect their ratings of students' BELAs was also important. Nine academic staff made up the academic staff sample, and questionnaire data was obtained from eight of the nine academic staff. The third group of participants were BELA users, who were other stakeholders using BELA results in their roles at Bond University. The BELA users sample was included as it was important to obtain insight regarding the consequences of BELA results. In total, 20 employees were identified as the target BELA user population with questionnaire data being obtained from 19 BELA users. Finally, two language assessment experts (i.e., External rater 1 and External rater 2) were included, as was necessary to gain objective insight from individuals not associated with the research site. These two participants assessed approximately 10% of the BELA scripts to analyse inter-rater reliability and provided feedback via semi-structured online interviews. Purposive sampling was used to recruit participants from each of the four stakeholder groups.

In terms of research tools, eight instruments were employed in the current study. These included: BELA, the Major Essay and overall grade in Core 1: Critical Thinking and Communication, as well as students' GPA after two semesters. Additionally, three questionnaires were utilised, including a questionnaire for students, a questionnaire for academics and a questionnaire for BELA users. Finally, the ASC feedback survey, which included two items regarding BELA, was included in this study. Analyses were conducted using IBM SPSS version 28 for quantitative data and NVivo 12 Pro for qualitative data.

Research question 1 (i.e., Can a satisfactory validity argument be presented for the Bond English Language Assessment [BELA]?) comprised five sub-questions, which reflected the inferences in Knoch and Elder's (2013) PELA validation framework. Here, sources of evidence to support asserted claims involved analysis of quantitative data, including student background data, academic performance data, questionnaire data (i.e., student, academic staff and BELA users questionnaires) and BELA reliability data. Qualitative data gathered via

questionnaires, feedback surveys and semi-structured interviews with external raters were also used to support claims made. Research question 2 (i.e., What factors predict academic performance of university students?) required statistical analysis (i.e., logical regression analysis). The focus here was on ascertaining predictions of students' academic performance (i.e., failing and high achievement). Appendix M summarises the responses to each research question, including the data collected and analyses conducted.

4.9 CONCLUSION

This chapter presented the methodology utilised in the current study. Specifically, it outlined the study's research aims and research questions, the overall research design, the sample and sampling strategy, the research tools used, the data analysis procedures and, finally, how each research question was answered. The next chapter presents the results of the analyses.

CHAPTER 5: RESULTS

5.0 INTRODUCTION

This thesis set out to present a validity argument for the Bond English Language Assessment (BELA) and, subsequently, achieve the overarching aim of determining what factors predicted academic performance of undergraduate students at a university. This chapter synthesises the results of analyses to respond to the study's research questions. The first research question (RQ1) posed whether a satisfactory validity argument could be presented for BELA. Responding to RQ1 required investigating five sub-questions, based on the five inferences of the validity framework. The second research question (RQ2) was whether BELA and/or other factors could predict academic performance of university students. For all statistical analyses, an alpha level of 0.05 was set, unless otherwise stated. The chapter is divided into two main sections; section 5.1 presents results concerning the validity argument for BELA (RQ1), and section 5.2 presents the results regarding academic performance predictors of undergraduate university students (RQ2). This chapter concludes with a summary of the study's results.

5.1 VALIDITY ARGUMENT FOR BELA

This section of the results chapter presents a validity argument for BELA. It focuses on responding to RQ1, whether a satisfactory validity argument can be presented for BELA, divided into five sub-questions. Section 5.1.1 presents the results of the Evaluation inference (RQ1a). Section 5.1.2 presents the results of the Generalisability inference (RQ1b). Section 5.1.3 presents the results of the Explanation and Extrapolation inference (RQ1c). Section 5.1.4 presents the results of the Decisions inference (RQ1d). Finally, section 5.1.5 presents the results of the Consequences inference (RQ1e). At the end of each section, a summary of results is presented.

5.1.1 EVALUATION INFERENCE

This section presents results obtained regarding how adequately BELA scores reflect the behaviours of test takers (i.e., undergraduate university students). Specifically, it responds to RQ1a, whether BELA scores are adequate reflections of observed behaviours. RQ1a is aligned with the Evaluation inference of Knoch and Elder's (2013) framework, which

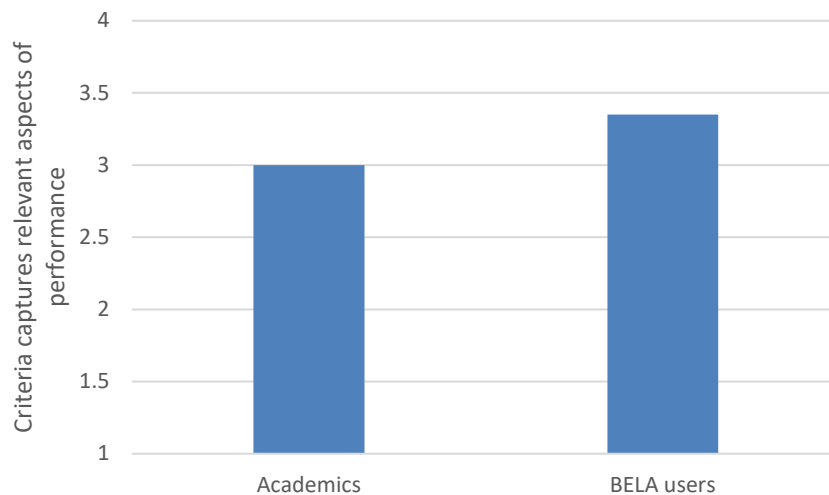
comprises five warrants. The next sections document the results concerning each warrant of the Evaluation inference.

5.1.1.1 WARRANT 1

Warrant 1 requires that the scoring criteria and rubrics capture relevant aspects of performance. To assess the relevance of the criteria comprising the BELA rubric, responses to the question, “The criteria (1. organisation, 2. linking and flow, and 3. grammar and vocabulary) capture aspects of performance most relevant to academic essay writing”, were analysed using a two tailed, independent samples t-test. Responses were obtained from academic staff ($n = 8$) and BELA users ($n = 17$), who indicated their level of agreement using a 4-point Likert scale from 1 (*Strongly disagree*) to 4 (*Strongly agree*). There was no significant difference in opinion regarding the criteria between academics ($M = 3.00$, $SD = 0.54$) and BELA users ($M = 3.35$; $SD = 0.49$), $t(23) = -1.63$, $p = .117$, indicating that both groups agreed the criteria used for BELA captured aspects of performance most relevant to academic writing (see Figure 5.1).

Figure 5.1

Mean comparison between academic staff and BELA users (N = 25) of the belief that the criteria used for BELA captured aspects of performance most relevant to academic writing



Furthermore, through semi-structured, online, one-to-one interviews with the researcher, external raters were asked to provide feedback on the rubric, including whether the criteria were relevant for assessing academic essay writing. The three criteria, Organisation, Linking and Flow, and Grammar and Vocabulary, were stated to be relevant for

academic essay writing (e.g., “I think they fit the purpose” and “I think for what it is, maybe these criteria work”). However, External rater 2 identified one aspect potentially missing from the rubric, which was either “task completion” or “task response”. The rater stated that it was important to assess whether students had responded to the task and presented an adequate argument, stating: “I was wanting to say something about whether or not the student even has answered the question and is on topic with their essay, and I was also looking for something about argument and whether the students made a logical argument.” In contrast, External rater 1 was not concerned with the absence of such a criterion. When asked if “task response” was required, the rater opined:

“I don't know. Does it really matter? ... I guess by including things like thesis – ‘contains a clear thesis statement’ that is kind of task response, I guess. Like if they [students] don't do that and you don't really know what their view is, ... according to these criteria, then they would be Below satisfactory with organisation.”

Warrant 1, that the BELA scoring criteria and rubric capture relevant aspects of academic essay writing, was supported. The feedback from one external rater regarding the lack of a specific task response criterion was noted and will be considered for future iterations of BELA. However, it is not considered imperative, based on overall feedback from stakeholders.

5.1.1.2 WARRANT 2

Warrant 2 necessitates that raters are able to implement consistent scoring procedures. Determining this involved analysing feedback from academic staff and external raters regarding their belief in their ability to apply the rubric via the academic staff questionnaire and semi-structured interviews respectively. Moreover, whether students believed their writing was correctly rated was determined via the student questionnaire. Finally, inter-rater reliability analysis was conducted using the ratings of the researcher (i.e., the second rater; $n = 180$, 16.3% of the sample), as well as two external raters (External rater 1, $n = 52$; External rater 2, $n = 68$, 10.8% of the sample) compared to the original ratings of academic staff.

Regarding beliefs about rating, academic staff were asked to indicate, using a 4-point Likert scale from 1 (*Strongly disagree*) to 4 (*Strongly agree*), whether they “can consistently apply the BELA criteria”. Of the eight respondents, the mean was 3.13 ($SD = 0.35$) with a mode of 3 (Agree; $n = 7$; 87.5%), suggesting that academics agreed they could consistently

apply the BELA criteria. Furthermore, students were asked to answer whether their “writing was correctly rated”, again using an identical 4-point Likert scale. Of the 99 student respondents, the mean was 3.20 ($SD = 0.61$) with a mode of 3 (Agree; $n = 65$; 65.7%). This indicated that, overall, students agreed that their writing was correctly rated.

External raters were also asked to provide feedback via semi-structured interviews on their ability to apply the rubric. The external raters expressed that, as it was a new system of rating writing in comparison to what they were used to (e.g., IELTS, placement tests), it was initially difficult to apply the rubric, but not impossible. Overall, the raters expressed that there were three main areas contributing to their difficulty in applying the rubric.

First of all, External rater 2 noted that they had some difficulty distinguishing between the first two criteria, Organisation and Linking and Flow, stating “I was a bit unsure about how to handle the difference between Organisation and Linking and Flow.” The rater further explained:

“I felt that that we were really looking at the same area twice. So for me that felt that there was some doubling up. So I felt that I was dealing with organisation, and then I was looking again at cohesion, and so I felt like I was really looking at that same aspect of a text twice, so that for me is something that I just found a bit unusual about the criteria.”

External rater 1 did not note an overlap between the first two criteria. In contrast, External rater 1 indicated they had difficulty with the third criterion, Grammar and Vocabulary. Specifically, the external rater expressed a preference to separate the two into standalone criteria. “In terms of Grammar and Vocabulary, I found that a bit tricky because I generally want to separate those out so that we can assess those things separately.” External rater 2, however, indicated they did not experience the same difficulty with rating Grammar and Vocabulary, stating:

“Grammar and Vocabulary was good because I felt OK ... I felt I was on familiar ground because I can think about grammar in terms of range, but also frequency of error and so accuracy and I could see the word choice. So that was great for vocab, word forms and the style.”

External rater 2 did recommend removing “punctuation” from the second criterion, Linking and Flow, and placing it within the third criteria, explaining, “for me, I wanted to put punctuation in Grammar and Vocabulary.”

The second challenge noted by both raters was not having a midpoint between “Below satisfactory” and “Satisfactory”. External rater 1 noted having difficulty with the absence of a midpoint in the 2-point rating scale. The rater explained, “there's only kind of one and three and nothing in between. I found it quite difficult there to make a choice between the two sometimes.” The rater elaborated:

“Just you know, for example like in grammar it says, ‘Frequent grammatical errors, limited range of vocabulary with frequent errors in Word choice or word formation’ So for me, it goes from ‘frequent grammatical errors’ to only ‘occasional errors’. That's a 3, with nothing in between, so I found it very difficult sometimes to assign [a rating].”

External rater 2 concurred regarding the lack of a midpoint on the scale, stating it was “challenging” to rate BELA essays with only a 2-point scale. The rater explained:

“Of course, you want to sit on the fence sometimes. Of course, you wanna say look, yes, there was some errors, but you know the student used a fantastic range ... but then you think well, OK, yes, the student had some really nice vocabulary or something or you know they used a really good range of structures but then you think yeah, but you know there were issues with run on sentences or you know, all their spelling was awful or punctuation [was problematic].”

Subsequently, External rater 2 explained that although there were times when they wanted to rate a criterion between “Below satisfactory” and “Satisfactory”, such as “needs work” or “partially achieved”, they realised that the 2-point scale forced them to make a decision. The rater reflected, “I want to sit on the fence. But I understood that I couldn't do that, and so I just had to [make a decision].” The rater later concluded that being forced to make a decision (i.e., not having a midpoint), was in fact beneficial, noting, “because you had to make that decision, that was probably a good thing.” Consequently, it is believed the challenge of not having a midpoint in the rating scale was overcome.

The third challenge identified, according to one external rater, was a lack of specificity in the rubric's descriptors. External rater 1 expressed:

“As I was marking them, I found it a little bit challenging to ... grade occasionally, because it was, it's kind of a binary system. There's like, not a lot of granularity there, so sometimes I found it hard to kind of look at descriptors in the criteria, which kind of accurately capture what I was kind of reading.”

The rater went on to recommend better defining the descriptors, arguing “it's just a matter of perhaps defining what some of your descriptors mean basically. ... If we look at grammar and vocabulary and occasional error in word choice there's different levels of error density aren't there? So what does an occasional error mean?” External rater 2, in contrast, did not indicate that the descriptors lacked detail.

Overall, both external raters acknowledged that their initial reactions to the rubric were primarily due to lack of experience applying it or “being a novice” rating BELA. External rater 1, for instance, initially explained, “I'm having a bit of trouble calibrating myself to it. Simply, I come from an IELTS background, so basically my background with assessment and moderating writing is from either IELTS or from bridging English [courses].” Furthermore, External rater 2 explained:

“Of course, using the IELTS criteria, yeah, for so long means that, that feels very familiar and comfortable for me. I'm on comfortable ground with that. So I don't want to say that I'm necessarily right. That's just the feeling I have when I mark. But having said all that, I was still able to use the [BELA] criteria and I think they fit the purpose.”

Additionally, External rater 2 acknowledged that the rubric was not flawed, just in contrast to the style of rubrics used previously. The rater stated, “I'm not at the point of saying ‘ohh it's flawed’, it's because I'm aware that I'm just coming from a different system.” Later, the raters both expressed a level of confidence in rating BELAs. For instance, External rater 2 described:

“After having my standardisation with you, Cameron, I was then able to feel that I was pretty confident to go in and assess. All of the standardisation materials were really helpful and it's really good to have those in your mind as a point of reference as

you're going through marking those papers and handy to be able to just go back and check to see what those moderated papers were like. So, you can always double check to check so that you feel you're staying on track, and I felt that once I got into it, I was able to do the marking fairly quickly and efficiently.”

Overall, both raters suggested that applying the criteria became easier over time. External rater 2 detailed, “I didn't feel like I was agonising over papers, not being able to decide where to put them. I felt like I was able to put them somewhere fairly confidently where I felt they probably deserve to be.”

Moreover, to assess inter-rater reliability, 180 students' BELA essays were selected by the researcher using systematic random sampling. These essays were rated by a second rater (the researcher) for analysis. As per the original BELA rating procedures, ratings were classified as 3 (Below satisfactory), 5 (Below satisfactory), 7 (Satisfactory), 9 (Satisfactory). Ratings were then analysed in comparison to the original ratings by academic staff ($N = 9$). The raters agreed regarding the marking of BELA for 101 of the 180 BELA rated. Cohen's kappa was calculated and found to reflect a just below fair level of inter-rater agreement ($K = .39, p < .001$). However, as BELA is used to screen which student may or may not have adequate academic writing skills, the same analysis was run but this time BELA scores of 3 and 5 was coded as Below satisfactory and a 7 or 9 was coded as Satisfactory. In this instance, the raters agreed regarding the rating of BELA for 143 of the 180 BELA rated. Cohen's kappa was calculated and found to reflect a fair level of inter-rater agreement ($K = .54, p < .001$).

To assess whether a specific marking criterion induced a higher level of inter-rater disagreement, it was also decided to assess the inter-rater reliability for the three criteria, Organisation, Linking and flow, and Grammar and vocabulary, all marked as 1 (*Below satisfactory*) or 3 (*Satisfactory*), respectively. For Organisation, the raters agreed for 157 out of the 180 BELAs rated. Cohen's kappa was calculated and found to reflect a good level of inter-rater agreement ($K = .65, p < .001$). For Linking and Flow, the raters agreed for 133 out of the 180 BELA rated. Cohen's kappa was calculated and found to reflect a fair level of inter-rater agreement ($K = .43, p < .001$). For Grammar, the raters agreed for 126 out of the 180 BELA rated. Cohen's kappa was calculated and found to reflect a fair level of inter-rater agreement ($K = .41, p < .001$). This meant there was a higher level of disagreement when assessing Grammar and Linking and Flow in comparison to Organisation.

Inter-rater reliability was also assessed for each individual academic ($N = 9$) who originally rated their students' BELA essays. Again, the Cohen's kappa was calculated for the BELA's original coding (3, 5, 7 and 9) and dummy coding (Below satisfactory; Satisfactory). While the coefficients were very high between some tutors and the 2nd rater (i.e., Tutor 5 and 7), there were substantial discrepancies with other tutors (i.e., Tutor 1, 4, 6 and 9), as shown in Table 5.1.

Table 5.1

Inter-rater reliability per tutor in the original coding and dummy coding ($n = 180$)

Tutor	2nd Rater	2nd Rater
	BELA Original Coding	BELA Dummy Coding
Tutor 1	.12	.18*
Tutor 2	.24*	.57*
Tutor 3	.45**	.57*
Tutor 4	.41**	.31
Tutor 5	.71***	.89***
Tutor 6	.36**	.50*
Tutor 7	.59***	.78***
Tutor 8	.26*	.60**
Tutor 9	.31**	.36*

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

In addition to the inter-reliability analysis between the original tutor ratings and ratings by the second rater (i.e., the researcher), as discussed above, inter-rater reliability was assessed using a sample of BELA essays rated by the two external raters, external rater 1 and external rater 2. Fleiss' kappa (Fleiss, 1971; Fleiss et al., 2003) was run to ascertain if there was agreement between the original overall BELA ratings of the nine tutors (i.e., scores of 3, 5, 7 or 9), the second rater (the researcher), and two external raters (i.e., external rater 1 and external rater 2). External rater 1 rated 68 BELAs that had been rated by the tutors and the second rater. Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019) amongst the three groups of raters' ratings, $\kappa = .06$ (95% CI, -.03 to .15), $p = .160$. When examining the ratings of external rater 1 and the tutors, Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019), $\kappa = .07$ (95% CI, -.078 to .23), $p = .337$. When looking at the ratings of

external rater 1 and the second rater (the researcher), Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019), $\kappa = .18$ (95% CI, .02 to .34), $p = .030$.

As the purpose of BELA is to screen students, the overall ratings were dummy coded, such that scores of 3 and 5 were coded as 1 (Below satisfactory) and scores of 7 and 9 were coded as 2 (Satisfactory). Fleiss' kappa was run again to determine agreement between the tutors, second rater and external rater 1. Here, Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019) amongst the three groups of raters' dummy coded ratings, $\kappa = .30$ (95% CI, .15 to .42), $p < .001$. Comparing external rater 1 and the tutors, Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019), $\kappa = .15$ (95% CI, -.09 to .39), $p = .207$. Comparing external rater 1 and the second rater (the researcher), Fleiss' kappa demonstrated fair agreement (Allen et al., 2019), $\kappa = .41$ (95% CI, .17 to .65), $p < .001$. See Table 5.2 below.

Table 5.2

Inter-rater reliability using Fleiss kappa (κ) per external rater 1 in the original coding and dummy coding (n = 68)

Rater comparison	Overall Original Coding, κ	<i>p</i> values	Confidence intervals	Overall Dummy Coding, κ	<i>p</i> values	Confidence intervals
Tutors/Rater 2/External rater 1	.06	.160	[-0.03; 0.15]	.30	< .001	[0.15; 0.42]
Tutors/External rater 1	.07	.337	[-0.08; 0.23]	.15	.207	[-0.09; 0.39]
Rater 2/External rater 1	.18	.030	[0.02; 0.34]	.41	< .001	[0.17; 0.65]

External rater 2 rated 52 BELAs that had been rated by the tutors and the second rater. Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019) amongst the three groups of raters' ratings, $\kappa = .27$ (95% CI, 0.18 to 0.37), $p < .001$. When examining the ratings of external rater 2 and the tutors, Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019), $\kappa = .19$ (95% CI, 0.03 to 0.35), $p = .022$. When looking at the ratings of external

rater 2 and the second rater (the researcher), Fleiss' kappa demonstrated fair agreement (Allen et al., 2019), $\kappa = .45$ (95% CI, 0.24 to 0.57), $p < .001$.

Overall ratings were again dummy coded for external rater 2 (i.e., scores of 3 and 5 were coded as 1 [Below satisfactory]; scores of 7 and 9 were coded as 2 [Satisfactory]). Fleiss' kappa was run to determine agreement between the three groups of raters. Here, Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019) amongst the three groups of raters' dummy coded ratings, $\kappa = .38$ (95% CI, .21 to .52), $p < .001$. Comparing external rater 2 and the tutors, Fleiss' kappa demonstrated below fair agreement (Allen et al., 2019), $\kappa = .13$ (95% CI, -0.14 to 0.41), $p = .336$. Finally, comparing external rater 2 and the second rater (the researcher), Fleiss' kappa demonstrated fair agreement (Allen et al., 2019), $\kappa = .54$ (95% CI, 0.27 to 0.81), $p < .001$. See Table 5.3 below.

Table 5.3

Inter-rater reliability using Fleiss kappa (κ) per external rater 2 in the original coding and dummy coding (n = 52)

Rater comparison	Overall Original Coding, κ	<i>p</i> values	Confidence intervals	Overall Dummy Coding, κ	<i>p</i> values	Confidence intervals
Tutors/Rater 2/External rater 2	.27	< .001	[0.18; 0.37]	.37	< .001	[0.21; 0.52]
Tutors/External rater 2	.19	.022	[0.03; 0.35]	.13	.336	[-0.14; 0.41]
Rater 2/External rater 2	.41	< .001	[0.24; 0.57]	.54	< .001	[0.27; 0.81]

Overall, insufficient support was demonstrated for Warrant 2, that raters can implement scoring procedures consistently. The lack of support was mainly due to the results of the inter-rater reliability analysis. Although raters believed they could rate BELA consistently, inter-rater reliability analysis indicated, in some cases, a low level of consistency.

5.1.1.3 WARRANT 3

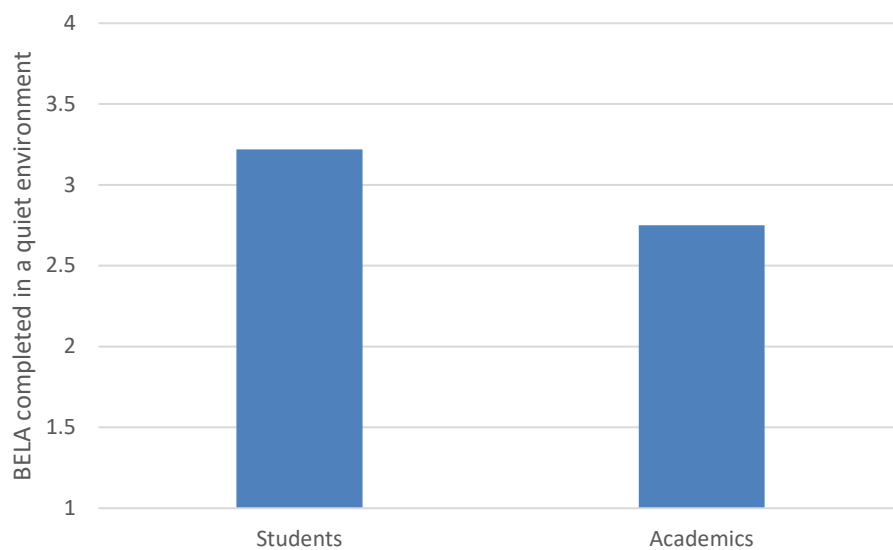
Warrant 3 states that test administration conditions are clear and appropriate. In response, students' reflection on the process they undertook in completing the test, as well as review of administration and procedures, was ascertained via the student questionnaire.

Perspectives of academic staff and BELA users via the relevant questionnaire and external raters' opinions via semi-structured interviews were also analysed.

To assess whether students completed the task in a quiet environment, the answers to the items, “I completed BELA in a quiet environment” for students ($n = 103$) and the item, “Students completed BELA in a quiet environment” for academic staff ($n = 8$) were compared using a two tailed, independent samples t-test. There was no significant difference in the mean scores between students ($M = 3.22$, $SD = 0.71$) and academics ($M = 2.75$; $SD = 0.71$), $t(109) = 1.81$, $p = .073$. The mode for students and academics was 3 (*Agree*; $n_{Students} = 49$, 47.6%; $n_{Academics} = 4$, 50%), suggesting that both students and academics mostly agreed that students completed the task in a quiet environment (see Figure 5.2).

Figure 5.2

Mean comparison of belief that students completed the task in a quiet environment between students and academics ($N = 111$)

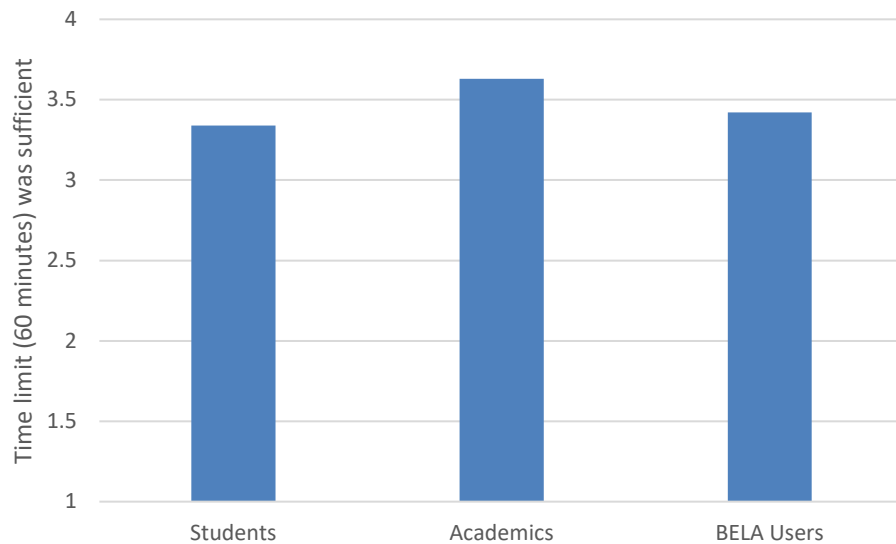


To determine whether there was enough time for students to complete BELA, the answers to the question “The time limit (60 minutes) is enough time to complete BELA” from the student questionnaire ($n = 103$), academic staff questionnaire ($n = 8$), and the BELA users questionnaire ($n = 19$) were compared using a one-way ANOVA. Here, the time limit being sufficient was the dependent variable (DV), while the three different questionnaires (i.e., students, academics, BELA users) were the independent variables (IVs). There was no significant difference in opinion between students ($M = 3.34$, $SD = 0.67$), academics ($M = 3.63$; $SD = 0.52$), and BELA users ($M = 3.42$; $SD = 0.51$), $F(2, 127) = 0.82$, $p = .444$. This

indicates that all groups of respondents believed that the time limit of 60 minutes was sufficient to complete BELA (see Figure 5.3).

Figure 5.3

Mean comparison of sufficient time limit (60 minutes) for students, academics and BELA users ($N = 130$)



Furthermore, the question of whether 60 minutes was sufficient to complete BELA was posed to the external raters. Both raters responded affirmatively, with External rater 1 stating “absolutely. Yeah. 60 minutes is a generous time allocation, I think” and then External rater 2 responding, “in terms of the time you've allowed them, I think it's adequate.”

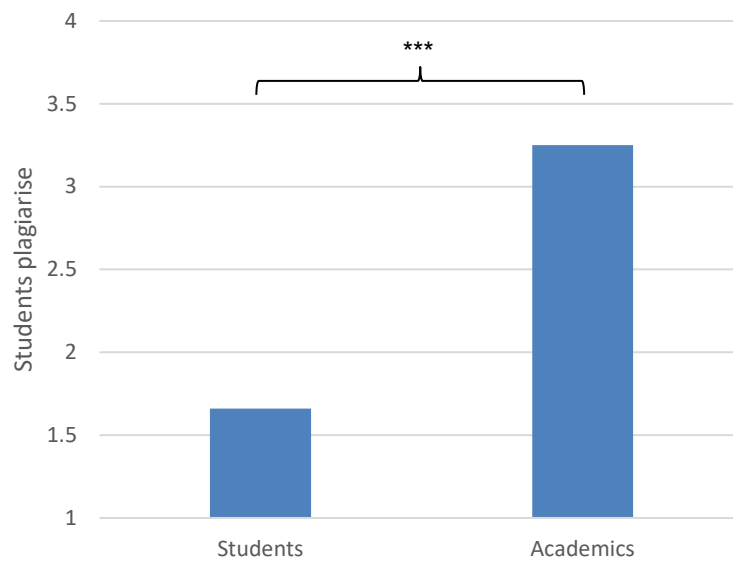
To determine whether students faced technical difficulties when completing BELA, the answers to the question “I had no technical difficulties when completing BELA” were analysed from the student questionnaire. Of the 103 responses, the mean was 3.51 ($SD = 0.62$) with a mode of 4 (*Strongly agree*; $n = 60$; 58.3%) demonstrating that, students strongly agreed that they had not faced any technical difficulties when completing the online assessment.

Furthermore, to ascertain whether academic integrity or plagiarism was an issue for BELA, answers to the question, “I got help from other sources or copied others’ work when completing BELA” from the student questionnaire ($n = 103$) and the item, “Students plagiarise their BELA” from the academic staff questionnaire ($n = 8$) were compared using a two tailed, independent samples t-test. There was a significant difference in the opinions of

students ($M = 1.66$, $SD = 0.94$) compared to academics ($M = 3.25$; $SD = 0.46$), $t(12.05) = -8.47$, $p < .001$. This can be considered a large effect size ($d = 1.73$; Cohen, 1988). Notably, this indicated that academics believed students had plagiarised when completing BELA, while students expressed that they had not (see Figure 5.4).

Figure 5.4

Mean comparison of belief that students plagiarise between students and academics (N = 111)



Note. *** $p < .001$.

To further assess whether academic staff viewed plagiarism as an issue, answers to the item in the academic staff questionnaire, “Because BELA is a low stakes assessment, the chances of students plagiarising is low”, were analysed. Of the eight responses, the mean was 1.75 ($SD = .71$) with a mode of 2 (*Disagree*; $n = 4$; 50%). This indicated that academics disagreed that the chance of students plagiarising was low due to the low-stakes nature of the task. Again, this suggests that academics view academic integrity as an issue, regardless of the low-stakes nature of BELA.

In response to their belief that academic integrity was a problem, academics were also asked whether they believed that “BELA should be completed in a computer lab in the presence of an invigilator”. Of the eight responses, there was a lack of clarity as to whether invigilation in a computer room was required; the mean response was between *Disagree* and *Agree* ($M = 2.63$; $SD = .74$) whilst the mode was 2 (*Disagree*; $n = 4$; 50%). This suggested

that academics were uncertain whether requiring students to complete BELA in a computer laboratory with an invigilator was necessary.

One of the external raters also raised concerns regarding test security and maintaining academic integrity. External rater 2 noted:

“My feeling whilst I was looking through some of the papers was that occasionally there may have been a student who, because of a fear of loss of face, may have done a little bit of copying and pasting. Well, you see chunks of text where you think that doesn't quite seem to fit with the rest of their essay. So unless you have a lockdown browser or something where you know, you force them to do the test without having access to the Internet, then I think you're always going to get that.”

In response to academic integrity concerns, it was recommended by External rater 2 that a third version of BELA be created to maintain academic integrity and test security. The rater explained:

“A possible thing to investigate is whether you could have one more [version] ... just to take away that little element of predictability. You know, that might be worth thinking about, but it, you know, that's just me coming from probably a slightly different context where those topics have to be constantly refreshed.”

Considering the above analyses, through student questionnaire responses, as well as feedback from academic staff, BELA users and the external raters, partial support was found for Warrant 3, that the test administration conditions of BELA are clearly articulated and appropriate. The major concern identified was academic integrity. Staff consistently acknowledged that academic integrity was a concern; however, students indicated they had not acted counter to academic integrity policies. Thus, further investigation is required.

5.1.1.4 WARRANT 4

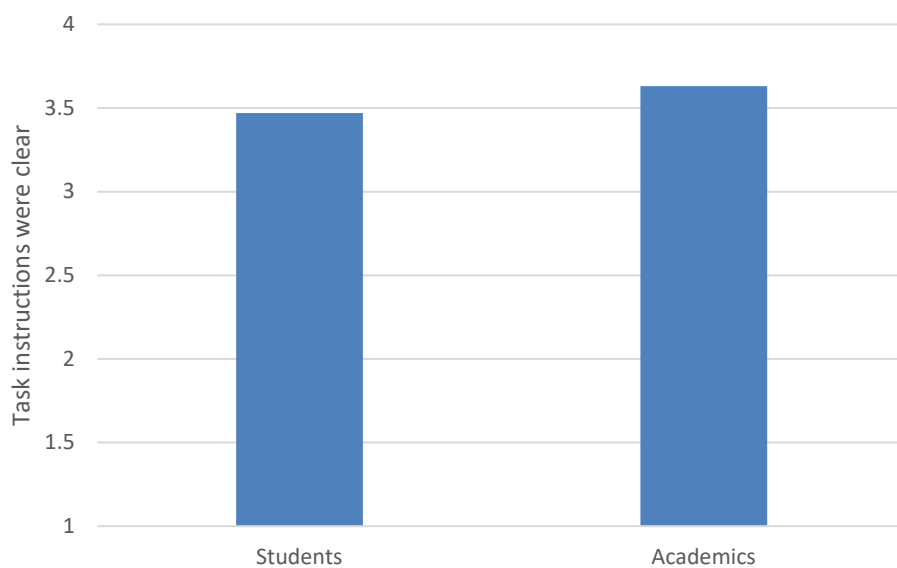
Warrant 4 requires clarity of both instructions and tasks. Responses in the student questionnaire and the academic staff questionnaire were analysed to determine whether the instructions and the overarching task for BELA were clear.

To assess whether test takers believed that the instruction and tasks were clear, the answers to the question “The instructions for BELA were clear” from the student

questionnaire ($n = 103$) and academics questionnaire ($n = 8$) were compared using a two tailed, independent samples t-test. There was no significant difference in scores concerning clear instructions when comparing the beliefs of students ($M = 3.47$, $SD = 0.68$) and academics ($M = 3.63$; $SD = 0.52$), $t(109) = -0.64$, $p = .522$. This demonstrates that both students and academics agreed that the instructions and tasks for BELA were clear (see Figure 5.5).

Figure 5.5

Mean opinion comparison of the clarity of the task and instructions for both students and academics (N = 111)



In addition, students ($n = 103$) were asked to respond to the item, “I understood the BELA question”. Overall, results indicated that students agreed with the statement ($M = 3.50$, $SD = 0.64$), with a mode of 4 (*Strongly agree*; $n = 57$; 55.3%), indicating that students strongly agreed that they understood the question.

Support was found through student and academic staff questionnaire responses for Warrant 4, that instructions and tasks for BELA are clear to all test takers.

5.1.1.5 WARRANT 5

Warrant 5 states that BELA should be appropriately pitched in terms of difficulty and is able to discriminate consistently based on students’ abilities. To assess test difficulty and discrimination power, it was first decided to dummy code the scores as Below satisfactory (scores of 3 and 5 = 0) and Satisfactory (scores of 7 and 9 = 1). Because two versions of

BELA exist, it was also decided to compute the test difficulty and discrimination separately. In terms of difficulty, both versions (i.e., version 1 and version 2) were above the ideal difficulty level of .75 but below the cut of .90 which would mean that the test was too easy (Haladyna, 1999). Moreover, both versions of BELA possessed a good level of discrimination $> .40$ (Kelley, 1939, see Table 5.4).

Table 5.4

Test difficulty and test discrimination scores according to each BELA version

BELA Version	Total No. of Students	Satisfactory Scores	Test difficulty	Test discrimination
Version 1	641	529	.83	.65
Version 2	506	427	.84	.58

Regarding test difficulty, external raters were asked whether they believed BELA was appropriate. Both raters confirmed that the assessment was appropriately difficult. External rater 1 commented, “I thought it was. I thought that both of the versions were appropriate in terms of difficulty level” and “they [the two versions] seem parallel and of equal difficulty.” Additionally, External rater 2 agreed that the versions were appropriately difficult, explaining:

“I thought they were manageable in terms of providing enough content for the students to write 300 to 400 words. So, there was enough for them to talk about, and they were relevant to their, to their own experiences and knowledge that they would have so that felt for me that they were the right level, and they were appropriate in terms of a stimulus, what they would, what students would be able to write. And I thought they were comparable in level as well.”

Overall, support was demonstrated for Warrant 5 that BELA was pitched appropriately in terms of difficulty and discriminates consistently based on students’ abilities. This was determined through statistical analyses of test difficulty and test discrimination, as well as feedback from external raters.

5.1.1.6 SUMMARY OF RQ1A

This section analysed the evidence gathered to respond to RQ1a, whether BELA scores are adequate reflections of observed behaviours, aligned with the Evaluation inference.

Table 5.5 details the type of evidence used to respond to RQ1a and whether each warrant was supported.

Table 5.5

Warrants and supporting evidence for the Evaluation inference for BELA

Evaluation inference		
RQ1a: Are BELA scores adequate reflections of observed behaviours?		
Warrants	Evidence	Supported?
1. Scoring criteria and rubrics capture relevant aspects of performance.	External rater feedback Academic staff questionnaire responses	Yes
2. Raters can implement scoring procedures consistently.	Inter-rater reliability analysis External rater feedback Academic staff questionnaire responses Student questionnaire responses	No
3. Test administration conditions are clearly articulated and appropriate.	Student questionnaire responses Academic staff questionnaire responses BELA users questionnaire responses	Partially supported
4. Instructions and tasks are clear to all test takers	Student questionnaire responses Academic staff questionnaire responses	Yes
5. Test is pitched at appropriate difficulty level and test tasks/items discriminate consistently between more and less able candidates.	Test difficulty and discrimination analyses External rater feedback	Yes

As shown in the table, support was demonstrated for warrants 1, 4, and 5. Partial support was found for Warrant 3, “test administration conditions are clearly articulated and appropriate”. The main discussion point here were concerns regarding academic integrity. Moreover, insufficient support was determined for Warrant 2, “raters can implement scoring

procedures consistently”. Inter-rater reliability analysis indicated a lack of adequate consistency between raters.

5.1.2 GENERALISABILITY INFERENCE

In this section, results are presented in terms of BELA’s ability to yield consistent results across the contexts of its uses. This refers to the generalisability of BELA’s two versions (version 1 and version 2). This section responds to RQ1b, whether BELA yields results consistent across assessment contexts. RQ1b is connected to the Generalisability inference, made up of four warrants. The following sections detail the results for each warrant.

5.1.2.1 WARRANT 1

Warrant 1 necessitates different versions of BELA being parallel in design. Responding to this warrant required comparison of the two versions, first of all, by the researcher and then objectively by the external raters.

A second version of BELA, version 2, was created due to academic integrity concerns. Rotating two versions of BELA each semester was considered appropriate to minimise the risk that students would plagiarise the work of students from a semester earlier. Furthermore, as students who fail Core 1: Critical Thinking and Communication likely repeat it the following semester, it was considered necessary to create a second version of the assessment. Both versions are education-based topics. Version 1 requires students to discuss whether face to face education is more effective than online classes (i.e., “Attending class is far more effective than studying online.” Discuss), whilst version 2 requires students to discuss whether exams are the best method of assessing students’ knowledge (i.e., “Exams are the best way to assess a student’s understanding of a university subject.” Discuss). It is posited that all undergraduate students would be familiar with both topics. The only difference between the two versions of BELA is the essay prompt. Hence, it is argued by the researcher that the two versions are indeed parallel in design.

External raters were asked whether the two versions of BELA were parallel. Both raters affirmed that versions 1 and 2 were similar in design. For instance, External rater 1 expressed, “they seem parallel”, whilst External rater 2 concurred, noting, “they were relevant to their, to their own experiences and knowledge that they would have And I thought they were comparable in level as well.” The rater also commented that both versions’

tasks contained “good topics.” This is consistent with the support found for Warrant 5 of the Evaluation inference.

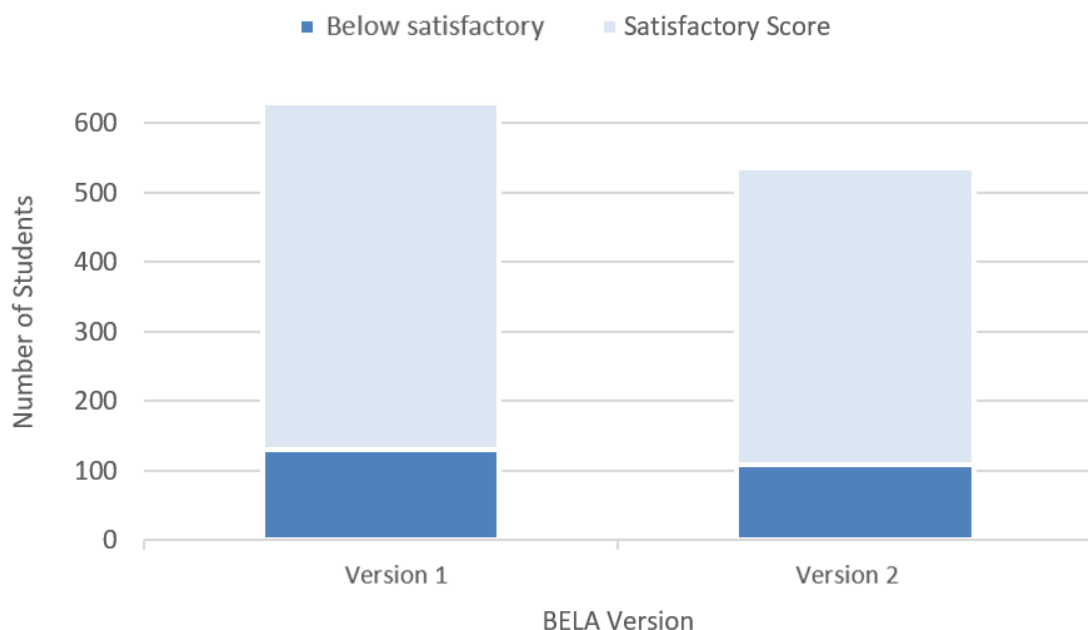
Warrant 1, that different versions of BELA were parallel in design, was supported. The two versions of BELA are highly similar. Objective review by external raters noted the versions were parallel in nature. The only difference between the two are the prompts.

5.1.2.2 WARRANT 2

Warrant 2 ensures that different forms of BELA comprise equivalent difficulty. This involved, first of all, the use of Pearson’s chi-square test of contingencies to determine if there was a difference in test results (i.e., Below satisfactory; Satisfactory) between the two versions of BELA. When examining the test difficulty score computed in response to Warrant 5 of the Evaluation inference, both scores were very similar (.83 and .84) for each BELA version. A Pearson’s chi-square test of contingencies was used to evaluate whether there was a difference in test results (Below satisfactory; Satisfactory) between the two versions of BELA. The chi-square test was not statistically significant, $\chi^2 = .07, p = .788$, implying that there was no statistically significant difference in scores between BELA version 1 and BELA version 2 (see Figure 5.6).

Figure 5.6

Bar chart illustrating the number of Below satisfactory versus Satisfactory students’ scores regarding BELA version (N = 1,164)



Moreover, external raters were asked for their insight regarding the difficulty of BELA. External rater 1 speculated that version 1, on the topic of online versus face to face education, may have been simpler, due to the pandemic and students' experience studying online, opining, "Well, I think in terms of the last two years, I guess. It's become very topical, I guess." The rater hypothesised that version 2, on the topic of exams "requires a bit more thought, I guess, to discuss in an intelligent way perhaps. I'm not sure." In contrast, External rater 2 responded that version 2, on the topic of exams, resulted in better essays. The rater argued:

"I personally thought the exams question probably produced a better essay. I don't know, that was just my feeling, that they had some quite interesting things to say about that. It allowed the good students to really pull out some really nice examples and some really nice concepts, whereas I felt with the online learning one it tended, in their responses, they tended to sort of focus more, a little bit more, on the social aspect as sometimes, you know, making friends."

External rater 2 later clarified that it may have just been due to personal preference in terms of what the rater preferred to read. The rater explained, "probably preferred reading the exam response, but maybe that's just because I'm, you know, a teacher and you know, I'm an assessment nerd!"

Overall, support was found for Warrant 2, different forms of BELA comprise equivalent difficulty, as demonstrated by the analysis of test difficulty. Although external examiners speculated that one topic may have been simpler or produced a better essay, feedback was not definitive.

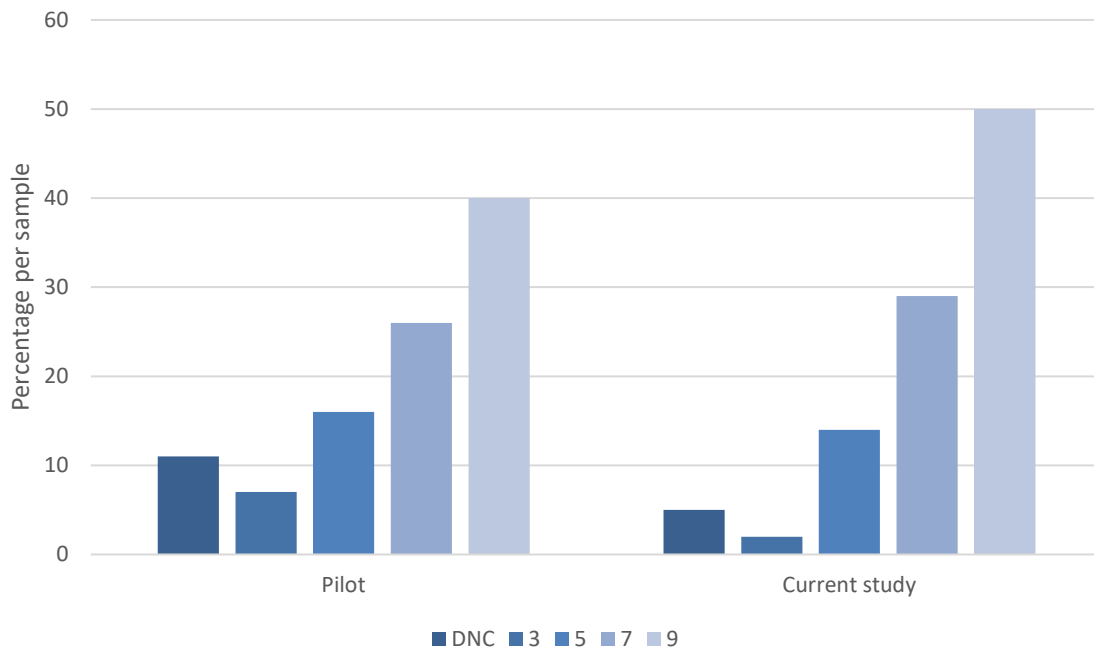
5.1.2.3 WARRANT 3

Warrant 3 focuses on sufficient tasks being included to provide stable estimates of students' essay writing skills. In order to determine whether the task provided stable estimates of task takers, pilot study results of a sample of 757 students conducted between 2014 and 2015 over three semesters were compared with the current study's data. A Pearson's chi-square test of contingencies was used to evaluate whether there was a difference in test results (i.e., Did not complete [DNC]; 3, 5; 7 and 9) between the pilot study ($n = 757$) and the current data ($n = 1,164$). The chi-square test was not statistically

significant, $\chi^2(4) = 46.44, p = .169$, indicating that BELA provides stable estimates of test taker ability (see Figure 5.7).

Figure 5.7

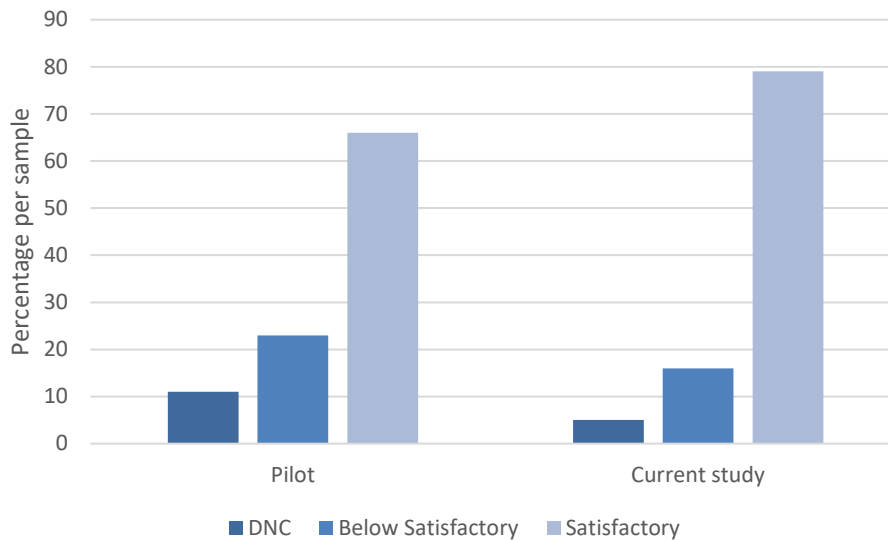
Bar chart illustrating the percentage per sample (pilot versus current study) of BELA results (i.e., Did not complete [DNC], 3, 5, 7 and 9; N = 1,921)



An additional Pearson's chi-square test of contingencies was used to evaluate whether there was a difference in test results between the pilot study ($n = 757$) and the current data ($n = 1,164$); however, this time, scores 3 and 5 were combined as Below satisfactory while 7 and 9 were deemed Satisfactory. The chi-square test was not statistically significant, $\chi^2(2) = 4.67, p = .097$, indicating that BELA provides stable estimates of test taker ability (see Figure 5.8).

Figure 5.8

Bar chart illustrating the percentage per sample (pilot versus current study) number of DNC, Below satisfactory and Satisfactory results (N = 1,921)



Warrant 3, that sufficient tasks are included in order to provide stable estimates of the abilities of test takers, was supported. As determined by the statistical analyses conducted, BELA provides stable estimates of students' essay writing skills.

5.1.2.4 WARRANT 4

Warrant 4 requires consistency of test administration conditions. Limited data was collected in response to this warrant, as BELA is completed outside of class time without the presence of an invigilator. However, findings regarding Warrant 3 of the Evaluation inference, indicated administration conditions were somewhat consistent. Students tended to complete BELA in a quiet environment, the time limit of 60 minutes was considered adequate and there were no technical difficulties encountered. Students strongly disagreed that they had received help from others or copied other students' work; however, concerns were raised by academic staff and one external rater regarding academic integrity. Thus, Warrant 4, that test administration conditions are consistent, was only partially supported. Limited data was obtained, and concerns raised regarding academic integrity necessitate further investigation.

5.1.2.5 SUMMARY OF RQ1B

This section presented analysis of the evidence gathered to respond to Research Question 1b, “Does BELA Yield Results Consistent Across Assessment Contexts (RQ1b)?”, which is aligned with the Generalisability inference. Table 5.6 summarises the warrants and evidence gathered to respond to RQ 1b and whether each warrant was supported.

Table 5.6

Warrants and evidence for the Generalisability inference for BELA

Generalisability inference		
RQ1b: Does BELA yield results consistent across assessment contexts?		
Warrants	Evidence	Supported?
1. Different test forms are parallel in design.	Researcher’s review of test forms External rater feedback	Yes
2. Appropriate equating procedures are used to ensure equivalent difficulty across test forms.	Pearson’s chi-square test of contingencies	Yes
3. Sufficient tasks are included to provide stable estimates of test taker ability.	Comparison of psychometric properties of the test as compared to trial population External rater feedback	Yes
4. Test administration conditions are consistent.	Academic staff questionnaire Student questionnaire responses	Partially supported

Overall, support was found for warrants 1, 2, and 3. Warrant 4, “Test administration conditions are consistent” is considered partially supported. The nature of BELA being an online, self-administered assessment means it is difficult to measure consistency. Stakeholders also voiced concerns regarding academic integrity.

5.1.3 EXPLANATION AND EXTRAPOLATION INFERENCE

This section includes the results obtained in response to two considerations. The first consideration was whether BELA provides information on students’ skills, knowledge and characteristics consistent with the understanding of academic ELP. The second consideration

was whether BELA is an adequate proxy for tasks that are performed in the domain of academia. This section presents results concerning RQ1c, which reflects the Explanation and Extrapolation inference, which consists of six warrants.

5.1.3.1 WARRANT 1

Warrant 1 requires determination of whether BELA results are adequate predictors of language performance at university. In order to estimate whether BELA test results (1 = Satisfactory; 0 = Below satisfactory) were good predictors of language performance (Major Essay; Overall Grade; GPA after two semesters), series of Point Biserial correlation were run. Students' whose BELA scores were satisfactory (i.e., scores of 7 and 9; coded as 1) significantly recorded higher Major Essay scores ($r_{pb}(1164) = .31, p < .001$), higher Overall Grades ($r_{pb}(1164) = .38, p < .001$), and higher GPA after two semesters ($r_{pb}(658) = .34, p < .001$) than their counterparts scoring Below satisfactory (scores of 3 and 5; coded as 0). Obtaining a satisfactory score on BELA explained 9.6%, 14.4%, and 11.6% of variance in Major Essay, Overall Grade and GPA after two semesters respectively. Consequently, Warrant 1, that BELA results adequately predict language performance at university, was supported.

5.1.3.2 WARRANT 2

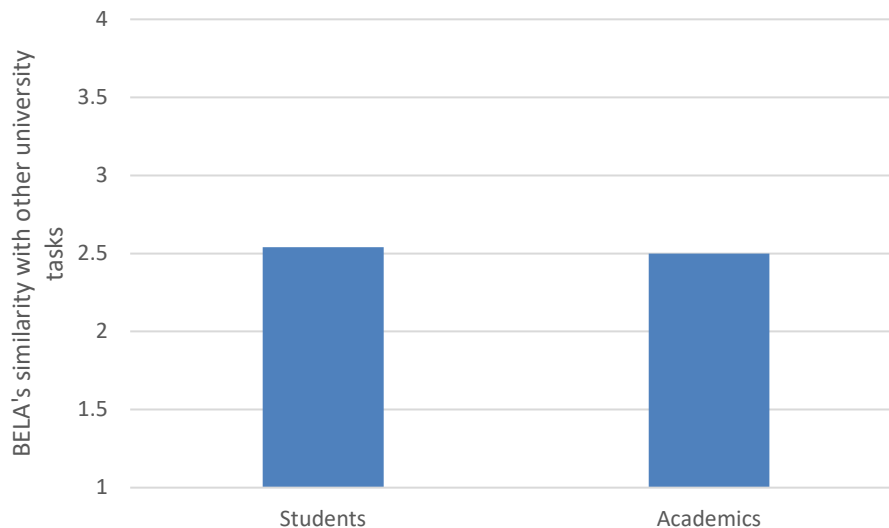
Warrant 2 necessitates similarity between BELA and tasks students are required to perform at university. This involved students, academic staff and BELA users reviewing BELA as to whether it captures relevant aspects of academic writing at university, as well as analysis of feedback from external raters and common types of assessment at the research site.

To assess whether stakeholders viewed BELA as being akin to tasks students are required to complete at university, responses to the question "BELA is similar to other university assessments tasks" from the students questionnaire ($n = 103$) and the academic staff questionnaire ($n = 8$) were compared using a two-tailed, independent samples t-test. There was a non-significant difference in mean scores between students ($M = 2.54, SD = 0.71$) and academics ($M = 2.50, SD = 0.54$), $t(109) = 0.17, p = .433$. For students, the mode was 3 (*Agree*; $n = 48, 46.6\%$) and for academics, multiple modes existed, 2 (*Disagree*; $n = 4, 50\%$) and 3 (*Agree*; $n = 4, 50\%$). This indicated that, amongst the two stakeholder groups,

there was a lack of consensus as to whether BELA was similar to other university assessment items (see Figure 5.9).

Figure 5.9

Mean opinion comparison of BELA's similarity with other university tasks for both students and academics (N = 111)



External raters were both asked whether the characteristics of BELA were similar to what is required in university assessment tasks. External rater 1 was initially uncertain as to whether BELA's characteristics were similar to university assessments. Overall, the rater expressed that BELA was an opportunity for students to demonstrate their writing skills, but was quite different to a university assignment, due to BELA's time limitation. The rater explained:

“This kind of routine task, it just shows their [students] ability to use the language and how proficient they are or how precise they are. I guess that's all it can do there's also the thing, of course, there's the time constraint as well. Like students, you know, if they're working, they're spending five hours on an essay as opposed to doing this quickly in 60 minutes. It's a whole different matter.”

However, External rater 2 affirmed that the characteristics of BELA were akin to university tasks. The rater noted that the requirements of BELA were similar to what was required of university assessment tasks with students having to respond to a question with

evidence. However, it was noted that research and referencing was not a requirement of BELA. When asked whether BELA was similar to other assessments, the rater explained:

“Yes, in that you have to present a response to a prompt, where you have to answer a question with your opinion. You have to support that opinion with some evidence. Of course, because it's a timed exam, you don't have the opportunity to go and find external sources to support your ideas that way. But you can still give examples from your own knowledge and reasoning from your own ideas, so yes, it's sort of a little microcosm of what people have to do at university.”

Moreover, External rater 2 expressed that requiring students to produce an academic essay, as opposed to another genre, was desirable. The rater maintained:

“An essay is the foundation of everything you write at university. It's just whatever we write at university, whatever genre we end up writing in, it's all based on that. If you can write a decent academic essay with paragraphs and with appropriate format, appropriate language and style, then you can turn your hand to most things at university.”

The genres of assessment used at the University were also investigated. The academic essay was identified as the second most common assessment genre, being included in 374 subjects (31.48%) at the time of the research. Essays were defined as “a short literary composition on a particular theme or subject, usually in prose and generally analytic, speculative, or interpretative” (Bond University, 2020, p. 2). Written reports were included in 422 subjects (35.52%); reflective essays were utilised in 90 subjects. Thus, essays and reflective essays appeared in 460 subjects (39.06%), making the essay the most common genre of assessment used.

Warrant 2, requiring similarity between BELA and academic tasks performed at university, was, overall, supported. Although academics were divided, external raters indicated that BELA and university assessments were aligned. The academic essay was also identified as one of the most frequently employed genres at the research site.

5.1.3.3 WARRANT 3

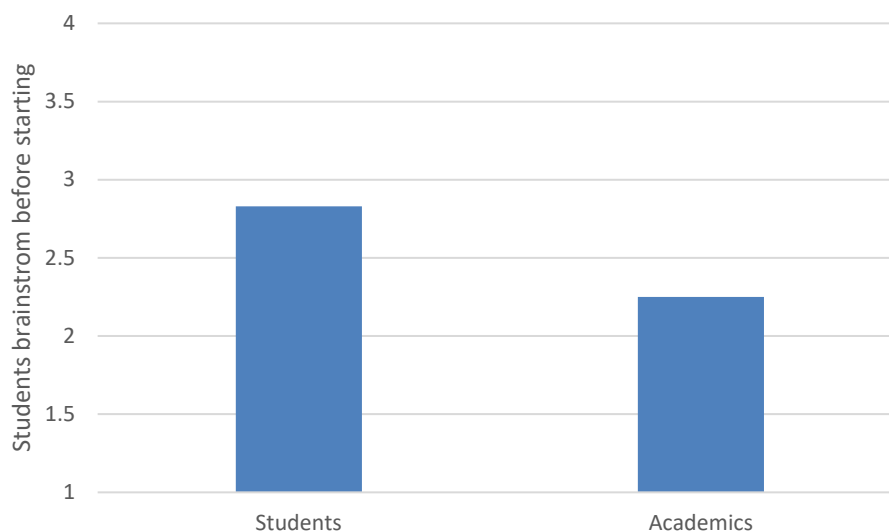
Warrant 3 requires consistency between what is required within academia and students' linguistic knowledge, processes and strategies utilised when completing BELA.

This involved analysis of student and academic staff responses to questionnaire items that focused on students' strategies employed when completing BELA and academic staff's opinions on whether language used by students when completing BELA was in line with linguistic expectations at university.

To determine the processes and strategies undertaken by students, in comparison to what is theoretically expected to occur, several items in the questionnaires were analysed. Firstly, whether students brainstormed ideas prior to commencing BELA was determined via the item, "I brainstormed ideas before starting my BELA essay" for students ($n = 103$) and "Students brainstorm ideas before starting their BELA essays" for academic staff ($n = 8$). Here, a two-tailed, independent samples t-test was run. There was a non-significant difference in mean scores between students ($M = 2.83$, $SD = 0.85$) and academics ($M = 2.25$, $SD = 0.71$), $t(109) = 1.87$, $p = .064$. For students, the mode was 3 (*Agree*; $n = 51$, 49.5%), and for academics, the mode was 2 (*Disagree*; $n = 4$, 50%). A definitive response was lacking amongst the two groups; however, students appeared to agree, on the whole, that they brainstormed ideas prior to starting the task (see Figure 5.10).

Figure 5.10

Mean opinion comparison of whether students brainstormed ideas before starting BELA for both students and academics (N = 111)

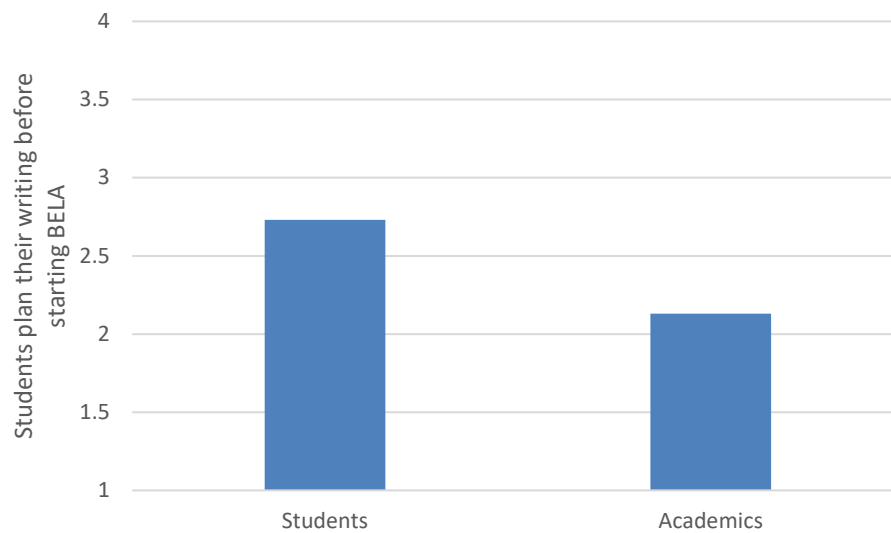


To gain further insight into the processes and strategies students undertook when completing BELA, whether students planned their writing was assessed by the item, "I planned my writing prior to starting the writing" for students ($n = 103$) and "Students plan

their writing prior to starting the writing” for academic staff ($n = 8$). A two-tailed, independent samples t-test was run, finding there was a non-significant difference in mean scores between students ($M = 2.73$, $SD = 0.87$) and academics ($M = 2.13$, $SD = 0.64$), $t(109) = 1.93$, $p = .057$. For students, the mode was 3 (*Agree*; $n = 46$, 44.7%) and for academics, the mode was 2 (*Disagree*; $n = 5$, 62.5%). A definitive response was lacking in terms of whether students planned their writing (see Figure 5.11).

Figure 5.11

Mean opinion comparison of whether students plan their writing before starting BELA for both students and academics (N = 111)



Furthermore, academic staff’s ($n = 8$) responses to the item, “The academic language students use in BELA is similar to the academic language used in other university tasks”, were analysed. Here, academic staff agreed with the statement ($M = 3.00$, $SD = 0.53$), with a mode of 3 (*Agree*; $n = 6$, 75%). This demonstrated that academic staff agreed that the language students used in BELA was akin to the language used in other university assessments.

Overall, Warrant 3, that students’ linguistic knowledge, processes and strategies utilised when completing BELA was consistent with the requirements of academic assessments, was partially supported. Some students expressed they had employed strategies when completing BELA, consistent with academic requirements, whilst others indicated they had not. Academics were equivocal in their responses concerning students employing strategies that were consistent with academic requirements, yet agreed that the linguistic

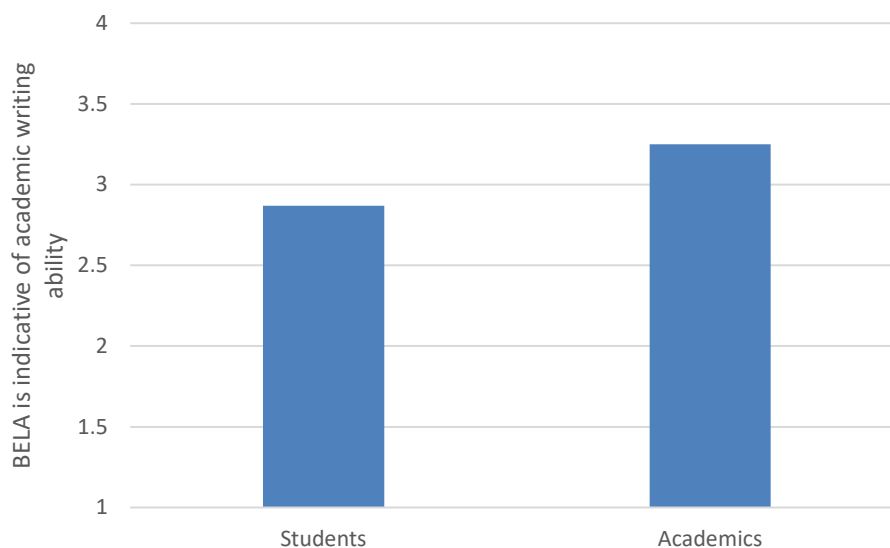
knowledge exemplified by students when completing BELA was consistent with academic expectations.

5.1.3.4 WARRANT 4

Warrant 4 concerns whether BELA scores provide sufficient information about students' academic language proficiency. To determine this, student ($n = 103$) responses to the item, "My writing in BELA was a good example of my academic writing ability" and academic staff's ($n = 8$) responses to the item, "A student's writing in BELA is a good example of his/her academic writing ability" were first analysed. Here, a two-tailed, independent samples t-test was run. There was a non-significant difference in mean scores between students ($M = 2.87$, $SD = 0.74$) and academics ($M = 3.25$; $SD = 0.46$), $t(109) = -1.42$, $p = .159$. For students and academics, the mode was 3 (*Agree*; $n_{Students} = 58$, 56.3%; $n_{Academics} = 6$, 75.0%), indicating that most students and academics, overall, agreed that BELA is representative of students' academic writing (see Figure 5.12).

Figure 5.12

Mean opinion comparison between students and academic staff's belief of BELA being representative of students' writing ability (N = 111)



Additionally, external raters were asked whether they felt BELA adequately assessed students' academic essay writing skills. External rater 1 was equivocal in their response regarding whether BELA adequately gauged students' academic essay writing skills. The rater pondered their own experience with creating assessments for different purposes (i.e.,

pre-entry assessments) and the challenge of achieving authenticity. When asked if BELA assessed students' writing skills, the rater explained:

“I don't know. It's a difficult question, I think. I really don't know But, you know, we do ponder that when we look at their writing, we do think to ourselves, will this student be able to handle, you know, the demands of academic writing based on this sample they've given us? And sometimes I don't know if we really know or not.”

When asked if BELA could be extended by requiring students to read an academic text and then paraphrase parts of the text to be incorporated in their own essay, External rater 1 noted, referring to the current BELA task, “I think it's fine”. Practicality constraints and resource limitations were cited as reasons for not extending BELA. External rater 1 maintained, “I think those things, they're just too hard to mark” and “it'll be way too time consuming.” External rater 1 did note that although BELA's authenticity was questionable, it was able to elicit “a good sample of their [students] use of language”. This enabled raters to determine “how proficient you are with grammar and just general writing. So I think it's fine. I don't think there's anything else you could really give them.”

On the other hand, External rater 2 maintained that BELA provided sufficient information about students' academic essay writing. They argued, “yeah, that's my feeling ... because it's not a gatekeeper test in that it's not deciding whether students are going to graduate from university. It's a test which is designed to assess whether or not students may require some support.” External rater 2 did acknowledge, however, that compromises were made to BELA due to practicality constraints. Ideally, BELA would provide additional information about students' ability to cite research, “it would be nice to ask students to write an extended essay and to give them, you know, the opportunity to support their ideas with sources.” However, it was explained that, due to resource limitations, it may not be practical: “unfortunately, you know, that logistically, that's really not possible, because your essays would be a lot longer and that would take people much more time to assess them.” The rater maintained that this was a common challenge faced by language assessment designers, “so basically, we have this dilemma, don't we when we try to assess academic writing? We have to balance the validity of the test with the practical considerations that we have.” The rater concluded, that in its current form, the compromises made regarding BELA were satisfactory. External rater 2 explained:

“So, for all those reasons, you have to sort of make compromises. And so you have to find a way where you can assess writing efficiently and in a cost effective way as well So, I feel that it's a compromise, but it's probably the best compromise.”

Consequently, Warrant 4, that BELA scores provided adequate information about students' academic language proficiency, was supported. Most students and academics agreed that BELA was indicative of students' academic writing skills. External raters maintained that due to practicality, BELA in its current form, provided adequate information about students' abilities.

5.1.3.5 WARRANT 5

Warrant 5 maintains that students' performance on BELA is related to their performance on other assessments of academic language proficiency. This warrant was not investigated. No data was available on students' scores in other measures of language proficiency, such as IELTS.

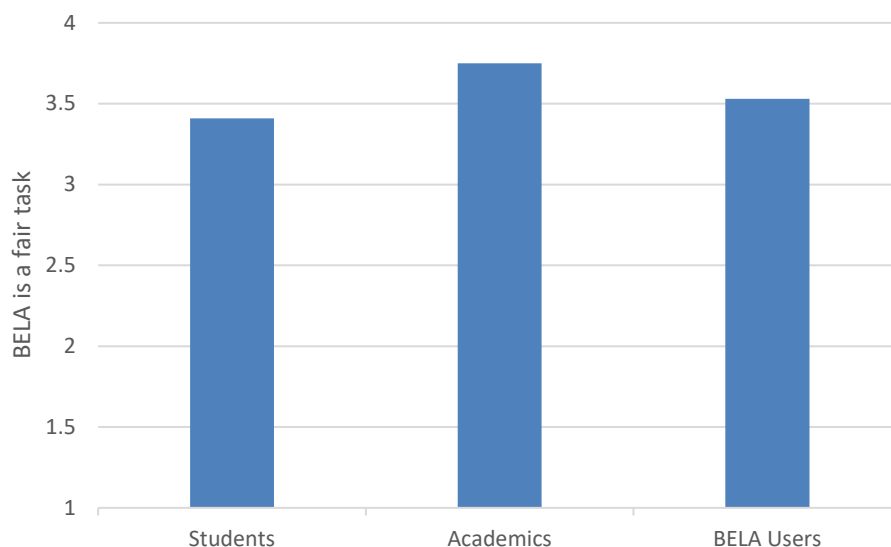
5.1.3.6 WARRANT 6

Warrant 6 necessitates that BELA does not unfairly favour certain groups of students. Feedback from students and academic staff was analysed via the student questionnaire and academic staff questionnaire, whilst external rater insight, through semi-structured interviews, was obtained. Statistical analysis was also conducted to compare the performance of English-speaking background students and EAL/D students.

First, to ascertain whether stakeholders believed BELA was fair, a one-way ANOVA was run to analyse responses to the item “BELA is a fair task” (DV) in response to the three different questionnaires (IVs; students, academics, BELA users). There was no significant difference between students ($M = 3.41$, $SD = 0.65$), academics ($M = 3.75$; $SD = 0.46$), and BELA users ($M = 3.53$; $SD = 0.51$), $F(2, 127) = 1.31$, $p = .274$. This indicates that each stakeholder group agreed BELA is a fair task (see Figure 5.13).

Figure 5.13

Mean opinion comparison of BELA as a fair task (N = 111)



To confirm students' perceptions of BELA as being fair, student responses ($n = 103$) to the item, "Some students have an unfair advantage over other students when completing the BELA task" were analysed. A definitive response from students was not obtained, yet most students expressed disagreement ($M = 2.25$, $SD = 0.78$), with a mode of 2 (*Disagree*; $n = 56$, 54.4%) that some had an unfair advantage when completing BELA.

Additionally, external raters were asked whether they perceived BELA was fair. External rater 1 indicated that it was difficult to determine whether BELA was in fact a fair assessment. They noted that what BELA required students to do was fair; however, fairness depended on the standards that were set for a satisfactory level of writing and whether students were aware of the standards. The rater described:

"What you're asking them [students] to do is fair. I'm still not sure so much if the criteria [is]. They don't know where the cut-off is, you know. I guess that's what I mean. I don't know where the cut-off is. What standard of grammar do you really, is really acceptable? Or not acceptable, I guess."

When asked if this constituted fairness, the rater was uncertain, indicating:

"I don't know if it is or not because, yeah, I don't know Would you say like there's this line and students who can write at that level or above, they don't need to do any

extra work. Students who are under or below this, they should or they need to or required to? Well, it depends where the line is.”

In contrast, External rater 2 indicated that they believed BELA was a fair measure of students’ writing. Specifically, when asked if BELA unfairly favoured certain groups of students, the rater described:

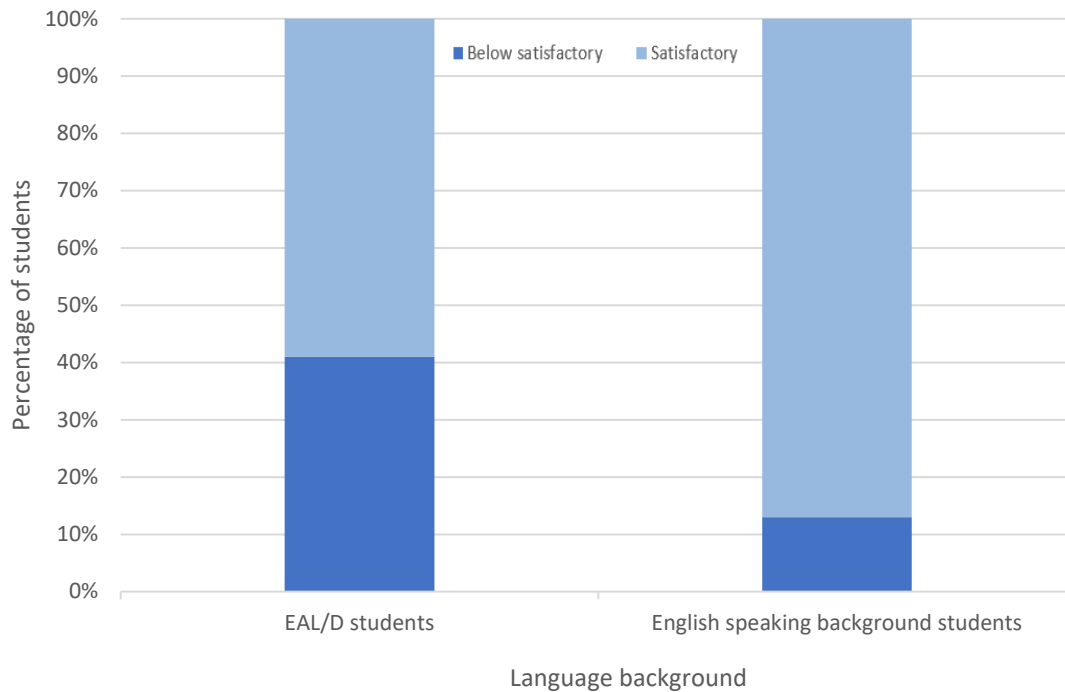
“In this day and age, it's pretty fair and pretty standard that you're asking people to do the test on online, but then everything's online nowadays, so you don't have that issue of people who can't use a computer or can't use a keyboard or anything like that. And so, I think that's fine in terms of being able to do it at home; they don't have to go into a test centre to do it. That's also convenient for the test takers. So, in terms of the time you've allowed them, I think it's adequate and also, probably just in terms of the security around it, you know, it's not because it's in a really high stakes test and there aren't sort of measures in place that would be overkill. Like, you know, remote proctoring or something like that. So, I think, yeah, I think it's suitable for the purpose of the test.”

Both raters, on the whole, expressed that what BELA required students to do was fair and did not unfairly provide advantage or disadvantage for any groups of students as test takers.

An additional consideration concerning Warrant 6 was language background and BELA performance. There is an expectation that EAL/D students would perform at a lower standard overall on BELA than English-speaking background students. In order to assess this statement, a Pearson’s chi-square test of contingencies was used to evaluate whether there was a difference in test results (Below satisfactory; Satisfactory) between language of origin (English speakers = 1; EAL/D = 2). The chi-square test was statistically significant, $\chi^2 = 107.17, p < .001$. In terms of percentage and to compensate for the non-equal sample size ($n_{\text{English-speaking background}} = 864; n_{\text{EAL/D}} = 300$), it becomes even clearer that approximately 41% of EAL/D students obtain Below satisfactory scores on BELA in comparison to 13.3% of English-speaking background students (see Figure 5.14).

Figure 5.14

Bar chart illustrating the percentage of Below satisfactory versus Satisfactory scores regarding language background (EAL/D; English-speaking background) (N = 1,164)



Warrant 6, that BELA tasks do not unfairly favour certain groups of test takers, was supported. Each stakeholder group strongly agreed that BELA was a fair task. A higher representation of EAL/D students in the Below satisfactory category was noted but expected.

5.1.3.7 SUMMARY OF RQ1C

This section analysed the evidence gathered to respond to Research Question 1c, whether BELA provides information on test takers' skills, knowledge and characteristics that keep within the understanding of academic ELP and is an adequate proxy for tasks performed in the academic domain. RQ1c aligns with the Explanation and Extrapolation inference. Support was found for warrants 1, 2, 4, and 6. Table 5.7 summarises the warrants and evidence sought to respond to research question RQ1c and whether each warrant was supported.

Table 5.7*Warrants and evidence for the Explanation and Extrapolation inference for BELA***Explanation and Extrapolation inference**

RQ1c: Does BELA provide information on test takers' skills, knowledge and characteristics that keep within the understanding of academic ELP? Is BELA an adequate proxy for tasks performed in the academic domain?

Warrants	Evidence	Supported?
1. Test results are good predictors of language performance in academic domain.	Point Biserial correlation analysis using BELA scores, scores on the Major Essay, overall grades in Core 1 and GPA after two semesters of study	Yes
2. Characteristics of test tasks are similar to those required of students in the academic domain (and those in the language development courses students are placed in).	Student questionnaire responses Academic staff questionnaire responses Assessment data at research site External rater feedback	Yes
3. Linguistic knowledge, processes, and strategies employed by test takers are in line with theoretically informed expectations and observations of what is required in the corresponding academic context.	Student questionnaire responses Academic staff questionnaire responses	Partially supported
4. Scores derived from the test provide sufficient information about candidates' academic language proficiency (i.e., no construct under-representation).	Academic staff questionnaire responses Student questionnaire responses External rater feedback	Yes

Explanation and Extrapolation inference

RQ1c: Does BELA provide information on test takers' skills, knowledge and characteristics that keep within the understanding of academic ELP? Is BELA an adequate proxy for tasks performed in the academic domain?

5. Performance on PELA relates to performance on other assessments of academic language proficiency.	Data not obtained	NA
6. Tasks do not unfairly favour certain groups of test takers.	Student questionnaire responses Academic questionnaire responses External rater feedback Statistical analysis of scores based on language background	Yes

As shown, partial support was found for Warrant 3, “Linguistic knowledge, processes, and strategies employed by test takers are in line with theoretically informed expectations and observations of what is required in the corresponding academic context”. No definitive answers in terms of whether students employed typical expected practices were ascertained. Furthermore, data was not obtained for Warrant 5 concerning performance on other language assessments.

5.1.4 DECISIONS INFERENCE

This section presents results obtained concerning how appropriate and well communicated decisions based on BELA scores are. It responds to RQ1d, whether the decisions based on BELA scores are appropriate and well communicated, reflecting the Decisions inference. The inference comprises six warrants, addressed in the next sections.

5.1.4.1 WARRANT 1

Warrant 1 states that students are correctly categorised based on their BELA scores. To determine this, academic outcomes of students were analysed. Firstly, the results on the Major Essay for Core 1: Critical Thinking and Communication for students scoring each of

the four possible scores were analysed. The same analysis was subsequently conducted regarding students' GPA after two semesters.

First, to investigate performance on the Major Essay, the results (scores out of 25) for students obtaining each of the four possible BELA scores (i.e., 3, 5, 7, 9) were compared. For students categorised as having Below satisfactory BELA scores, students who scored 3 ($n = 28$) on BELA, on average, scored lower on the Major Essay ($M = 8.23$, $SD = 6.75$) in comparison to students who scored higher on BELA. Further, students scoring 5 on BELA ($n = 163$), on average, scored higher ($M = 12.74$, $SD = 5.78$) than those scoring 3, yet lower than those scoring 7 or 9 on BELA. Regarding students categorised as having Satisfactory BELA scores, those scoring 7 ($n = 339$), on average scored higher on their Major Essay ($M = 14.13$, $SD = 6.21$) than students scoring 3 or 5 on BELA, whilst students scoring 9 on BELA ($n = 586$), on average, obtained the highest Major Essay results ($M = 16.53$, $SD = 5.06$). Figures 5.15 to 5.18 illustrate the distribution of Major Essay grades for each BELA score. Notably, students obtaining Below satisfactory BELA scores were slightly negatively skewed (3 = -1.184 and 5 = -0.893), whilst students obtaining Satisfactory BELA scores were more negatively skewed (7 = -1.087 and 9 = -1.774).

Figure 5.15

Histogram illustrating distribution of Major Essay results for students scoring 3 on BELA ($N = 28$)

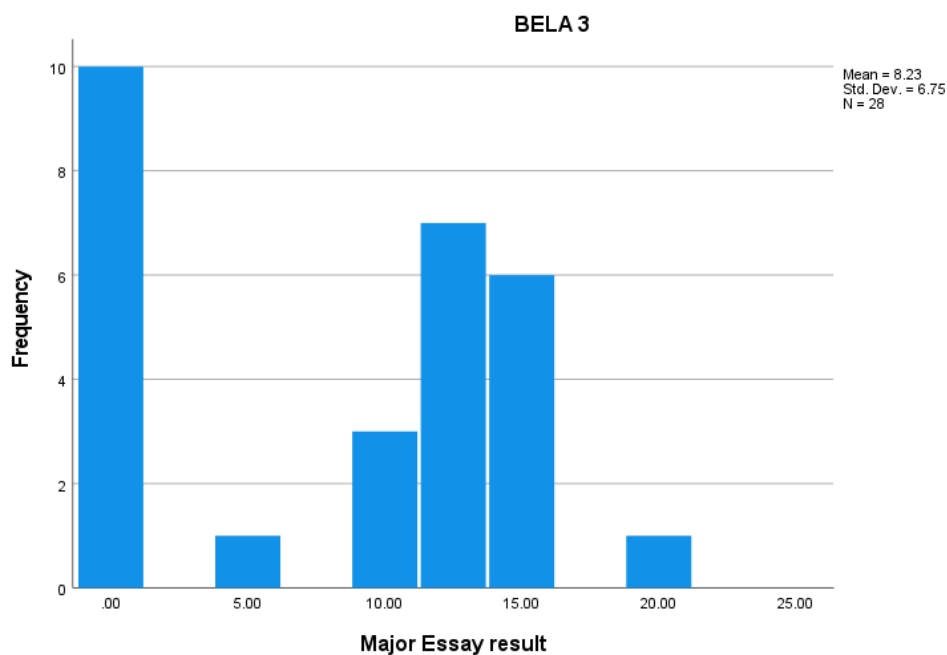


Figure 5.16

Histogram illustrating distribution of Major Essay results for students scoring 5 on BELA (N = 163)

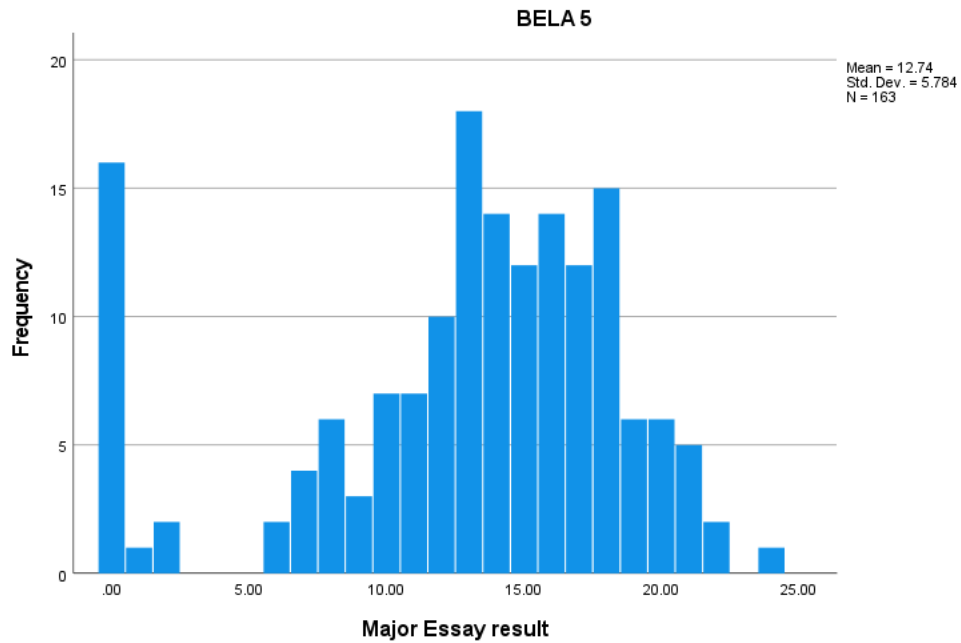


Figure 5.17

Histogram illustrating distribution of Major Essay results for students scoring 7 on BELA (N = 339)

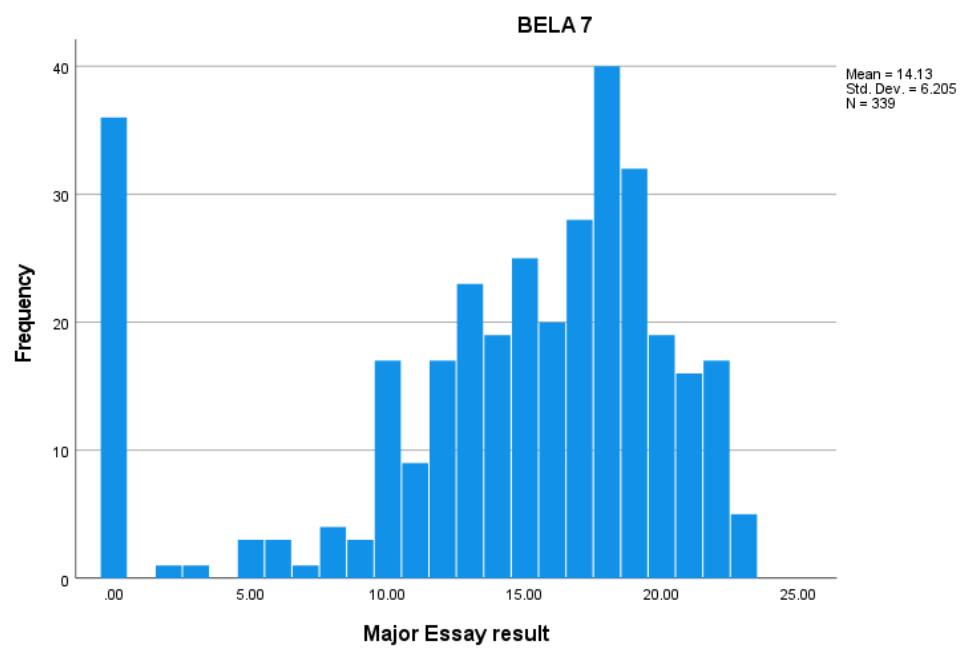
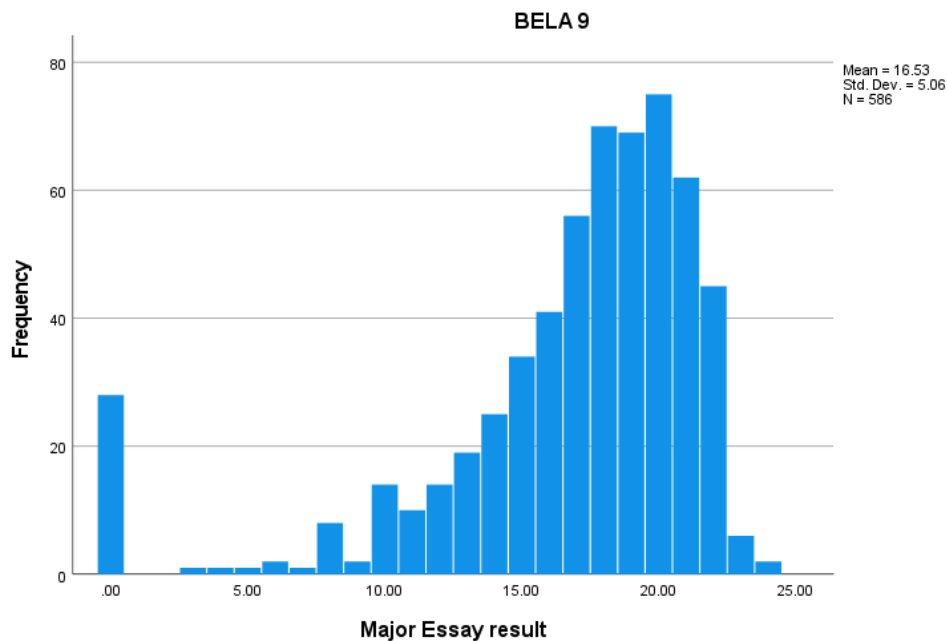


Figure 5.18

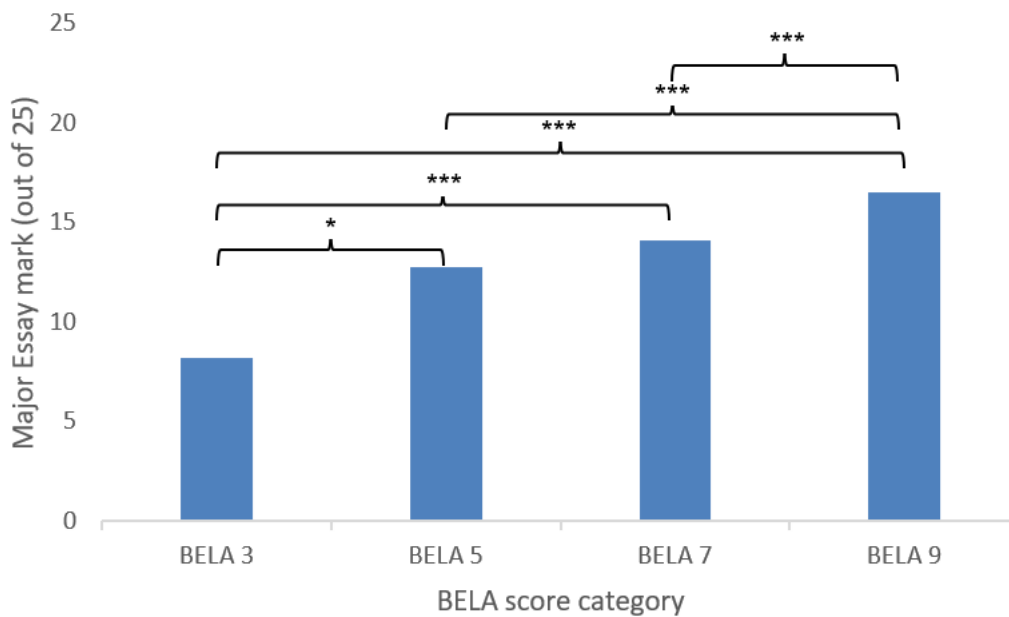
Histogram illustrating distribution of Major Essay results for students scoring 9 on BELA (N = 586)



Moreover, to determine if there were significant differences between students' scores on the Major Essay based on their BELA score category (i.e., scores of 3, 5, 7 or 9), a one-way ANOVA was run with Major Essay marks as the DV and BELA score categories as the IV. The Levene's statistic indicated that there was a violation of the assumption of homogeneity of variance ($p < .001$). The ANOVA was statistically significant, indicating that there was a difference in Major Essay marks regarding students' prior BELA scores, $F(3, 1112) = 40.05, p < .001$. This can be considered a medium effect size ($f = .30$; Cohen, 1988). Post Hoc analysis using Games-Howell revealed that students who scored 3 on BELA ($M = 8.23$; $SD = 6.75$) had significantly lower Major Essay marks than students who scored 5 ($M = 12.74$; $SD = 5.78$), 7 ($M = 14.13$; $SD = 6.20$) and 9 ($M = 16.53$; $SD = 5.06$). Additionally, students who scored 5 and 7 received significantly lower marks on their Major Essays than those who scored 9. However, there was no significant difference in Major Essay marks between students who scored 5 and 7 on BELA (see Figure 5.19).

Figure 5.19

Mean marks on Major Essay for each BELA score category



Note. * $p < .05$, *** $p < .001$

Second, GPA (0 - .99 = Fail; 1.00 – 4.00 = Pass) after two semesters was analysed for students obtaining each of the four scores. For students categorised as having Below satisfactory BELA scores, students who scored 3 ($n = 28$) on BELA, on average, maintained a GPA of below 1.0, indicating a fail grade ($M = .89$, $SD = .736$) in comparison to students scoring higher on BELA. Students scoring 5 on BELA ($n = 163$), on average, maintained a GPA above 1.0 ($M = 1.54$, $SD = .857$); however, their GPA was below those scoring 7 or 9 on BELA. Regarding students with Satisfactory BELA scores, those scoring 7 ($n = 339$), on average, maintained a slightly higher GPA ($M = 1.87$, $SD = 1.027$) than students scoring 5, whilst students scoring 9 ($n = 585$), on average, maintained the highest GPA ($M = 2.35$, $SD = .924$) equivalent to a Credit average. Figures 5.20 to 5.23 illustrate the distribution of GPAs for each BELA score. Notably, students receiving Below satisfactory BELA scores were positively skewed (3 = .933 and 5 = .390), whilst students obtaining Satisfactory BELA scores were negatively skewed (7 = -.011 and 9 = -.316). Please note, the dotted line in the below figures indicates GPA of 1.0 (i.e., Pass).

Figure 5.20

Histogram illustrating distribution of GPA results after two semesters for students scoring 3 on BELA (N = 28)

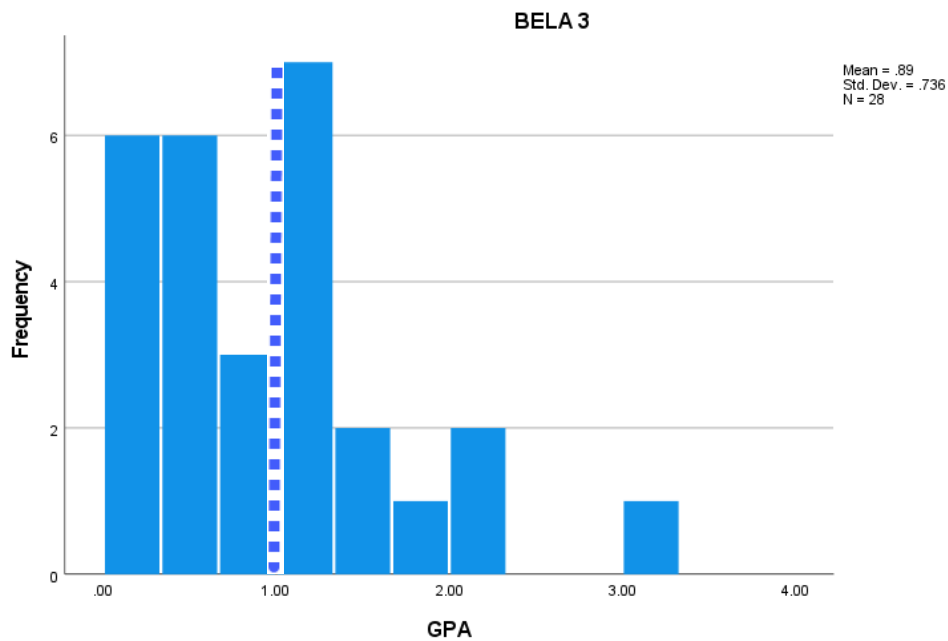


Figure 5.21

Histogram illustrating distribution of GPA results after two semesters for students scoring 5 on BELA (N = 163)

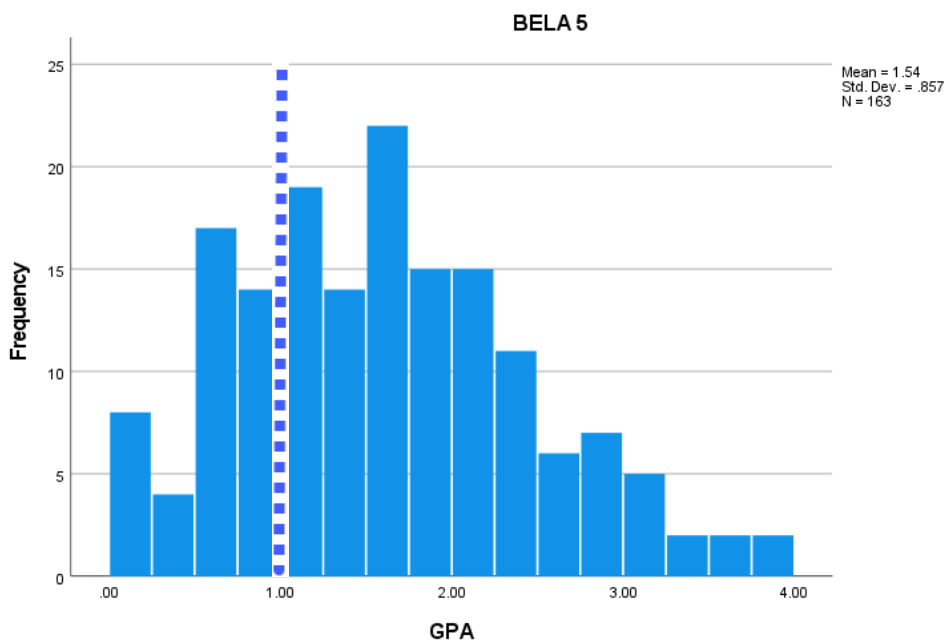


Figure 5.22

Histogram illustrating distribution of GPA results after two semesters for students scoring 7 on BELA (N = 339)

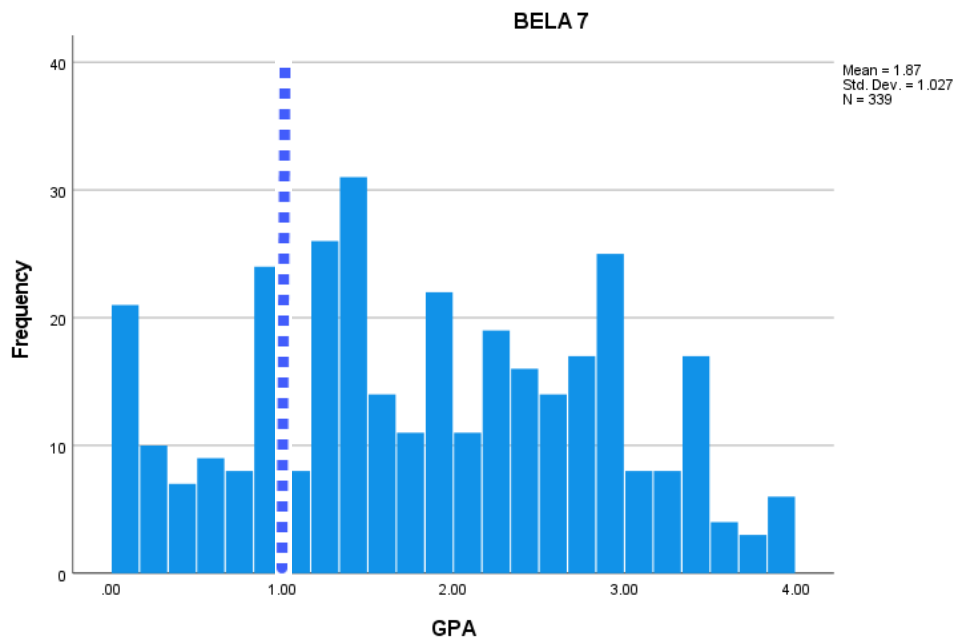
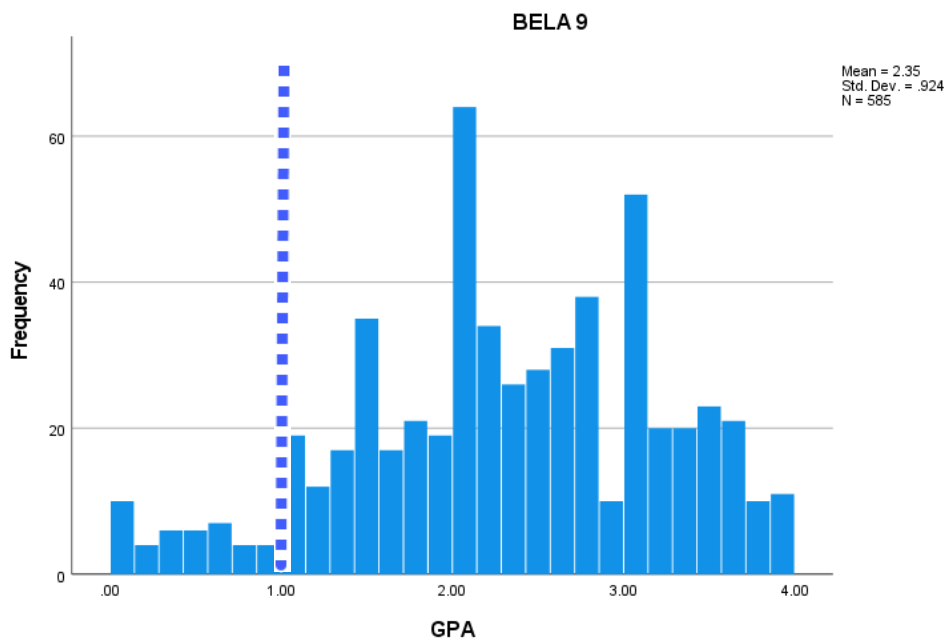


Figure 5.23

Histogram illustrating distribution of GPA results after two semesters for students scoring 9 on BELA (N = 585)

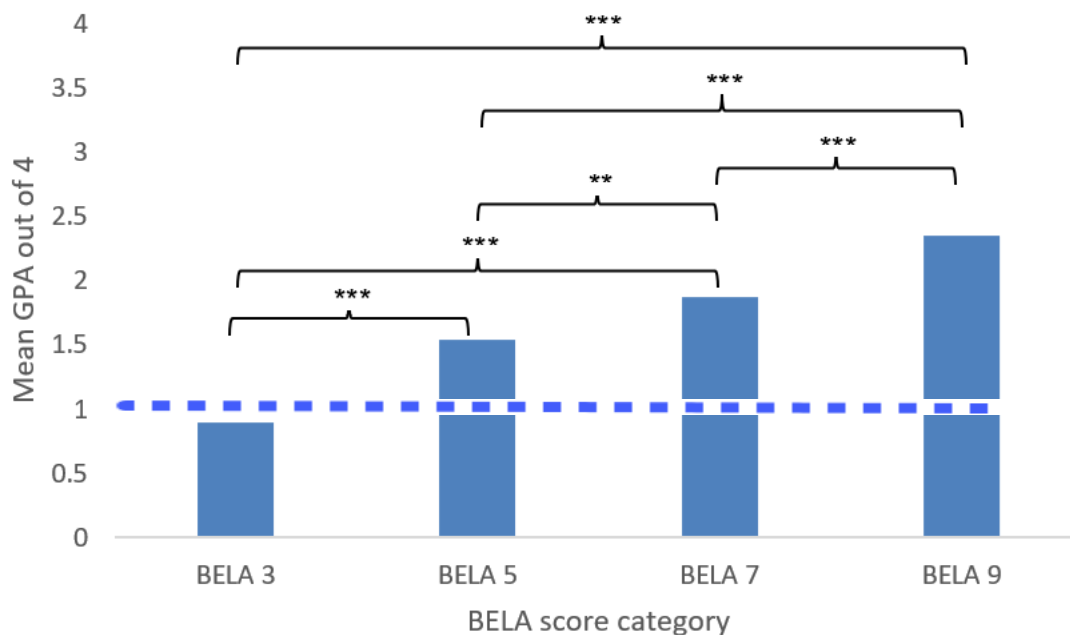


To evaluate whether there were significant differences between students' GPAs after two semesters based on their BELA score category (i.e., scores of 3, 5, 7 or 9), a one-way

ANOVA was run with GPA (out of 4.0) as the DV and BELA score categories as the IV. The Levene’s statistic indicated that there was a violation of the assumption of homogeneity of variance ($p < .001$). The ANOVA was statistically significant, indicating there was a difference in GPA regarding students’ prior BELA scores, $F(3, 1111) = 54.09, p < .001$. This can be considered a large effect size ($f = .40$; Cohen, 1988). Post Hoc analysis using Games-Howell revealed that students who scores 3 on BELA ($M = 0.89$; $SD = 0.74$) had significantly lower GPA after two semesters than students who scored 5 ($M = 1.54$; $SD = 0.86$), 7 ($M = 1.87$; $SD = 1.03$) and 9 ($M = 2.35$; $SD = 0.92$). Furthermore, students who scored 5 and 7 obtained significantly lower GPA after two semesters compared to those who scored 9. There was also a significant difference in GPA after two semesters between students who scored 5 and 7 in BELA. Figure 5.24 below illustrates this result; please note, the dotted line indicates GPA of 1.0 (i.e., Pass).

Figure 5.24

Mean GPA for each BELA score category



Note. ** $p < .01$, *** $p < .001$

Overall, the analysis comparing BELA performance and subsequent Major Essay results, as well as GPA after two semesters, provides evidence to support Warrant 1, that students are correctly categorised based on BELA results.

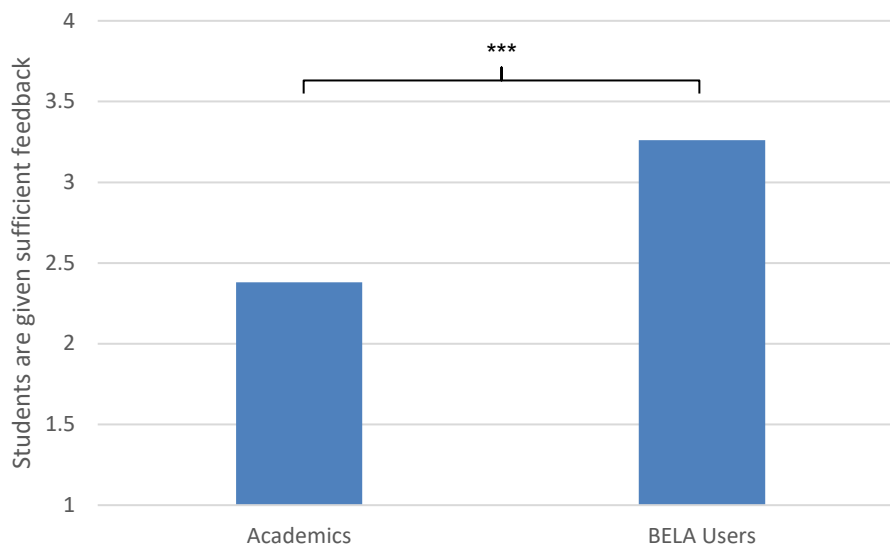
5.1.4.2 WARRANT 2

Warrant 2 necessitates that feedback on BELA performance is provided to students, as well as a recommendation based on results. Here, relevant items in the student questionnaire, academic questionnaire and BELA users questionnaire were analysed, in addition to qualitative feedback from students via the ASC feedback survey was also used to determine whether this warrant was supported.

First, to assess whether students received sufficient feedback, the answers to the question “Students are given sufficient feedback on their academic writing as part of the BELA process” from the academic staff questionnaire ($n = 8$) and the BELA users questionnaire ($n = 19$) were compared using a two-tailed, independent samples t-test. There was a significant difference in mean scores between academic staff ($M = 2.38$, $SD = 0.52$) and BELA users ($M = 3.26$; $SD = 0.65$), $t(25) = -3.41$, $p = .002$. This can be considered a large effect size ($d = 1.43$; Cohen, 1988). The mode for academics was 2 (*Disagree*; $n = 5$, 62.5%), whilst for BELA users, the mode was 3 (*Agree*; $n = 10$, 52.6%). This suggests there was uncertainty amongst academics on whether sufficient feedback was provided to students; however, BELA users felt sufficient feedback was given to students (see Figure 5.25).

Figure 5.25

Mean comparison of sufficient feedback for students between academic staff and BELA users (N = 27)



Note. *** $p < .001$.

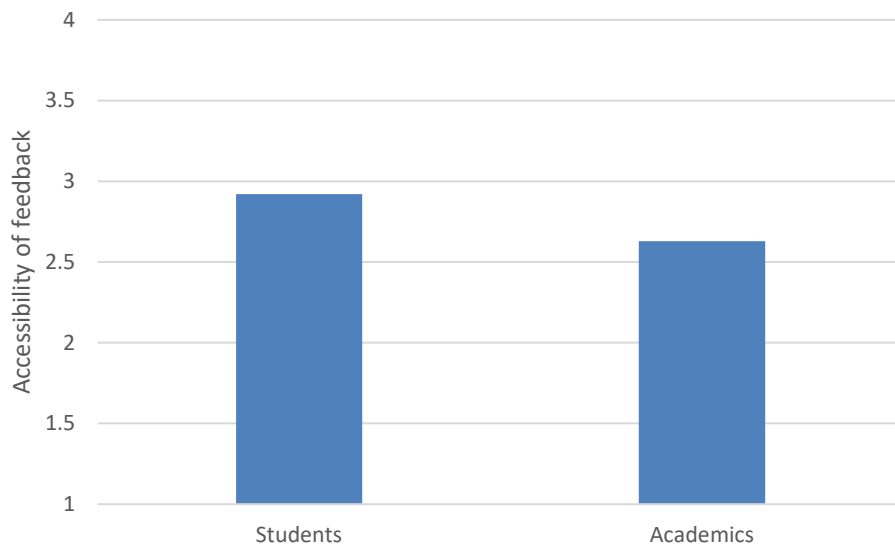
Qualitative responses from the ASC feedback survey were also analysed. Several students expressed that they desired feedback. One student noted, “Could have been beneficial to receive feedback.” Another student stated it was straightforward, but would have appreciated feedback, stating “It was self-explanatory. It would have been useful to know what I should have improved on.” Finally, one student indicated that more detailed instructions could have been provided, as well as feedback maintaining, “More details regarding instructions would have been nice (in terms of what was being assessed). Feedback would have been good too.” This data suggests that some students were not provided feedback.

Furthermore, in the qualitative, open-ended item in the students questionnaire, one student expressed, “There was no marking for BELA as it was marked out of 2 if you completed it. We received no feedback.” This again highlights a lack of feedback amongst some students.

Second, to assess feedback accessibility, the responses to the items, “It was easy to access feedback on my BELA” for students ($n = 99$) and “It is easy for students to access feedback on their BELAs” ($n = 8$) for academics were compared using a two-tailed, independent samples t-test. There was a non-significant difference in mean scores between students ($M = 2.92$, $SD = 0.53$) and academics ($M = 2.63$; $SD = 0.52$), $t(105) = 0.96$, $p = .340$. The mode for students and academics was 3 (*Agree*; $n_{Students} = 48$, 48.5%; $n_{Academics} = 5$, 62.5%). A definitive response to the question of whether feedback was easily accessible for students was lacking (see Figure 5.26).

Figure 5.26

Mean comparison of feedback accessibility between students and academics (N = 107)



Third, to determine whether students received a recommendation on the basis of their results, student responses ($n = 99$) to the item, “I received a recommendation/recommendations based on my BELA” were analysed. There was a lack of certainty concerning whether students believed they received a recommendation based on their BELA ($M = 2.44$, $SD = 0.96$) with a mode of 2 ($n = 34$, 34.3%).

Overall, a lack of support was established for Warrant 2, that BELA results include feedback on student performance and a recommendation. Triangulating various data types reveals that some students received little or nil feedback.

5.1.4.3 WARRANT 3

Warrant 3 requires recommendations to be linked to on-campus support. Of the 190 students who were flagged as having Below satisfactory academic essay writing, 87.34% ($n = 166$) attended ASC. Students were not directly asked whether BELA recommendations were linked to on-campus support. However, when students ($n = 99$) were asked whether they believed that “Academic writing support is available at the University”, they strongly agreed with the statement ($M = 3.59$, $SD = 0.54$). In contrast, BELA users’ responses ($n = 19$) to the item, “Recommendations provided to students according to their BELA scores are linked to the University's on-campus support”, reflected a strong belief that students were closely directed to on-campus support ($M = 3.58$, $SD = 0.51$). Hence, Warrant 3 was partially supported. Students clearly agreed that writing support was available on-campus, and BELA

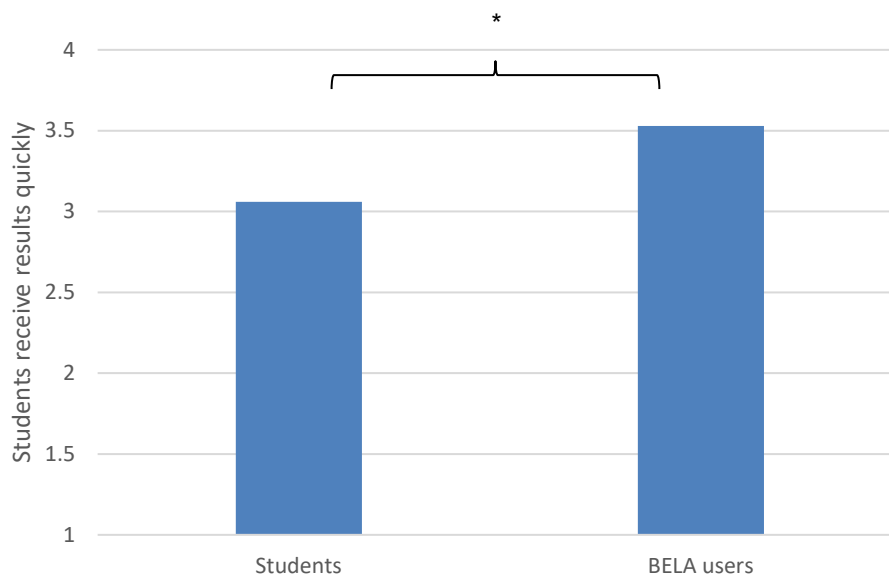
users indicated that recommendations were linked to on-campus support. However, considering the lack of support found for Warrant 2, further research is required to investigate all aspects of feedback, including the provision of feedback, recommendations and the connection between recommendations and available support on campus.

5.1.4.4 WARRANT 4

Warrant 4 requires BELA results to be delivered to stakeholders in a timely manner. Students are to complete BELA by Wednesday of week 2; academic staff rate students' writing prior to Monday week 3; and results are distributed on Monday of week 3 to students and BELA users. Stakeholder beliefs were ascertained by analysing responses to the student questionnaire and BELA users questionnaire. The answers to the question, "I received the results on my BELA quickly" from the student questionnaire and "BELA results are communicated to me in a timely manner" from BELA users questionnaire were compared using a two-tailed, independent samples t-test. There was a significant difference in mean scores between students ($M = 3.06, SD = 0.73$) and BELA users ($M = 3.53; SD = 0.70$), $t(116) = -2.58, p = .011$. This can be considered a medium effect size ($d = .66$; Cohen, 1988). Both students and BELA users agreed that the results were distributed in a timely manner, but with a significantly stronger positive opinion amongst BELA users (see Figure 5.27).

Figure 5.27

Mean comparison of timely feedback for students and BELA users (N = 118)



*Note. * $p < .05$*

Warrant 4, that BELA results be distributed in a timely manner, was supported, based on the opinions of both students and BELA users.

5.1.4.5 WARRANT 5

Warrant 5 suggests that BELA results should be available to all stakeholders. To determine this, responses to the item, “All university stakeholders who can influence students' retention are provided with BELA data” were analysed. BELA users ($n = 17$) agreed that all stakeholders were given access to BELA data ($M = 3.24$, $SD = 0.66$). Thus, Warrant 5 was supported.

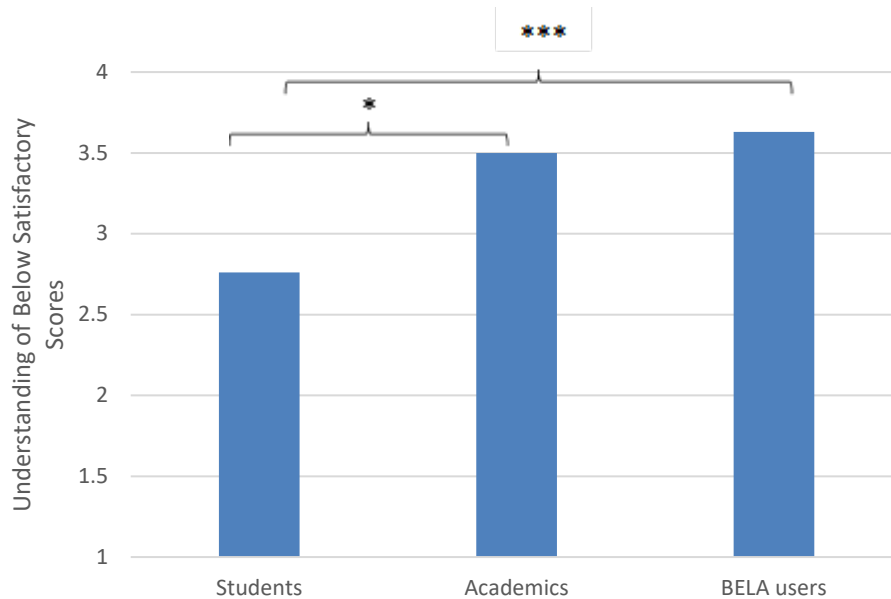
5.1.4.6 WARRANT 6

Warrant 6 requires that all BELA stakeholders understand the meaning and intended uses of BELA scores. This involved analysing the feedback from all BELA stakeholder groups via the three questionnaires.

To first determine whether users understood the meaning of BELA scores of 3 and 5 being below satisfactory, a one-way ANOVA was run to analyse responses to the item, “A score of 3 or 5 indicates that a student's academic writing is below satisfactory” (DV) amongst the three questionnaires (IVs; students, academics, BELA users). There was a significant difference between students ($M = 2.76$, $SD = 0.77$) with a mode of 3 (*Agree*; $n = 50$, 50.5%), academic staff ($M = 3.50$; $SD = 0.54$) with multiple modes existing 3 (*Agree*; $n = 4$, 50%) and 4 (*Strongly agree*; $n = 4$, 50%) and BELA users ($M = 3.63$; $SD = 0.50$), with a mode of 4 (*Strongly agree*; $n = 12$, 63.2%), $F(2, 123) = 14.17$, $p < .001$. This can be considered a large effect size ($f = .56$; Cohen, 1988). After post-hoc analysis, there was a significant difference between students compared to academics and BELA users. Staff stakeholders demonstrated clear understanding of the lower scores' meaning; however, results indicated that some students did not understand the meaning of scores 3 or 5 (see Figure 5.28).

Figure 5.28

Bar chart illustrating the mean of agreement to the meaning of 3 and 5 as below satisfactory ($N = 126$)

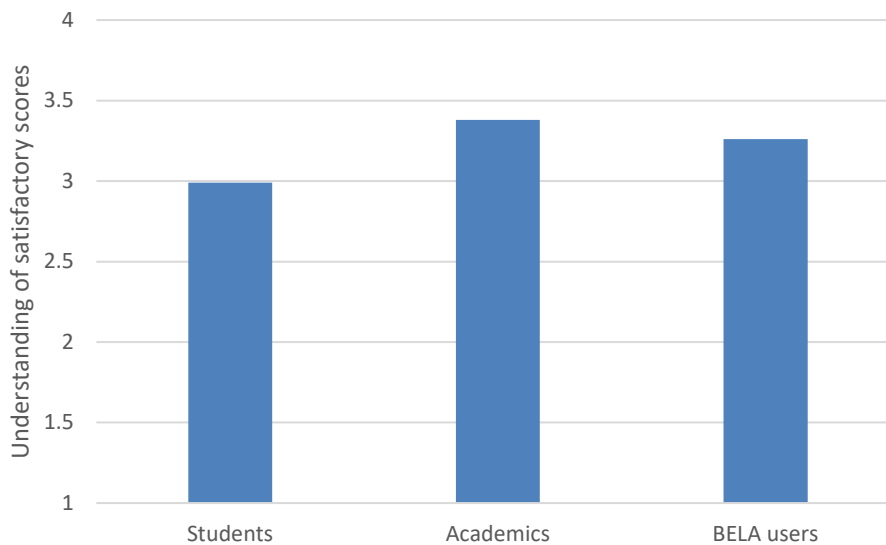


Note. * $p < .05$, *** $p < .001$.

Secondly, to see whether users understood the meaning of BELA scores of 7 and 9 being satisfactory, a one-way ANOVA was run concerning the answers to the question, “A score of 7 or 9 indicates that a student's academic writing is satisfactory” (DV) amongst the three different questionnaires (IVs; students, academics, BELA users). There was no significant difference between students ($M = 2.99$, $SD = 0.63$), academics ($M = 3.38$; $SD = 0.52$), and BELA users ($M = 3.26$; $SD = 0.56$), $F(2, 123) = 2.72$, $p = .070$. This demonstrated that most stakeholders understood that a score of 7 or 9 indicated satisfactory academic essay writing (see Figure 5.29).

Figure 5.29

Bar chart illustrating the mean of agreement to the meaning of 7 and 9 as satisfactory ($N = 126$)



To further assess students' understanding of the meaning of BELA scores, responses to the item, "If I didn't receive an email from my lecturer or tutor, it meant my writing was satisfactory", were analysed. Students ($n = 99$) indicated agreement that if they did not receive an email from their lecturer, it meant their academic writing was satisfactory ($M = 3.11$, $SD = 0.74$) with a mode of 3 ($n = 59$, 59.6%). This indicated that students believed if their lecturer or tutor had not emailed them regarding their performance on BELA, their writing was Satisfactory. Moreover, the item, "I do not know if my writing was satisfactory or unsatisfactory" was analysed. Students ($n = 99$) mostly disagreed with the statement ($M = 2.06$, $SD = 0.89$) with a mode of 2 (*Disagree*; $n = 42$, 42.4%), suggesting that the majority of students understood whether their writing, as assessed by BELA, was acceptable for university.

Additionally, within the qualitative, open-ended items in the students questionnaire, multiple students noted a lack of clarity regarding the scoring of BELA. One student explained, "I don't believe I ever received a score for my BELA writing task, so I was unable to determine in the questionnaire which scores indicated a satisfactory or not-satisfactory outcome." Another student expressed, regarding their responses to the questionnaire, "answers regarding BELA feedback may be inaccurate (sic) since I was not able to see any commentary or know where to find it. The only thing I could see is my score out of 2. So I was really confused about the 1-10 scaling referred to in the survey." One student

recommended changes to the scoring of BELA, opining, “I would suggest that the grading range is larger for each section.” This is notable, as it is consistent with indicative comments by external raters regarding the rubric not containing a midpoint.

Overall, Warrant 6, that all BELA stakeholders understand the meaning and intended uses of BELA scores, was only partially supported. Academic staff and BELA users indicated that they understood the meaning of Below satisfactory (i.e., scores of 3 and 5) and Satisfactory (scores of 7 and 9). In contrast, students expressed some uncertainty, particularly concerning Below satisfactory scores.

5.1.4.7 SUMMARY OF RQ1D

This section detailed the evidence gathered to respond to Research Question 1d, whether decisions based on BELA scores are appropriate and well communicated, connected to the Decisions inference. Table 5.8 summarises the warrants and evidence sought to respond to research question RQ1 and whether each warrant was supported.

Table 5.8

Warrants and evidence for the Decisions inference for BELA

Decisions inference		
RQ1d: Are decisions based on BELA scores appropriate and well communicated?		
Warrants	Evidence	Supported?
1. Students are correctly categorised based on their BELA scores.	Analysis of Major Essay results and GPA based on BELA scores	Yes
2. The test results include feedback on test performance and a recommendation.	Student questionnaire responses Academic staff questionnaire responses BELA users questionnaire responses ASC feedback survey	No
3. The recommendation is closely linked to on-campus support.	Student questionnaire responses BELA users questionnaire responses	Partially supported

Decisions inference

RQ1d: Are decisions based on BELA scores appropriate and well communicated?

4. Assessment results are distributed in a timely manner.	Student questionnaire responses BELA users questionnaire responses	Yes
5. The test results are available to all relevant stakeholders.	BELA users questionnaire responses	Yes
6. Test users understand the meaning and intended use of the scores.	Student questionnaire responses Academic staff questionnaire responses BELA users questionnaire responses	Partially supported

As illustrated in the table, support was found for warrants 1, 4, and 5. Partial support was determined for Warrant 3, “The recommendation is closely linked to on-campus support”. Students maintained that academic writing support was available on campus yet had indicated they had not received a recommendation. Similarly, warrant 6, “Test users understand the meaning and intended use of the scores” was partially supported. Students expressed a lack of clarity, in particular, regarding Below satisfactory scores. Moreover, a lack of support was evident for Warrant 2, “The test results include feedback on test performance and a recommendation.” Notably, numerous students indicated there was a lack of feedback provided (e.g., “we received no feedback”).

5.1.5 CONSEQUENCES INFERENCE

This section presents results obtained regarding the consequences of utilising BELA and whether the decisions based on results are of benefit to stakeholders. It responds to RQ1e, whether the consequences of using BELA and the decisions informed are beneficial to all stakeholders, aligned with the Decisions inference. The inference comprises eight warrants; the next sections document the results of each.

6.1.5.1 WARRANT 1

Warrant 1 requires that all targeted students complete BELA. Determining this involved analysis of completion data obtained via the Core 1: Critical Thinking and

Communication learning management system (i.e., Blackboard) over five semesters. Of the 1,176 students who were enrolled in the subject between May, 2018 and January 2020 semesters, 94.90% ($N = 1,116$) of commencing undergraduate students completed the assessment, which is a very high percentage of PELA completion, but still not universal uptake.

To determine reasons for non-completion, respondents ($n = 7$) to the student questionnaire who did not complete BELA were asked to indicate a reason for non-completion. The most common reason was late enrolment or delayed commencement in the subject. Indicative quotes included, “I was not still here when BELA 1 was held”, “I could not attend the first few weeks” and “I was sick recovering from surgery”. Other reasons provided included forgetting to complete the task (e.g., “I forgot that”) and not being able to access the task (e.g., “I couldn’t find the task on ilearn [sic]”).

In an effort to achieve universal uptake, one BELA user suggested, via questionnaire, to make the task a condition of enrolment, outside Core 1: Critical Thinking and Communication. The BELA user expressed:

“I’d also think that we could compel students to complete it as a condition of enrolment, rather than as a part of, CORE 1. Perhaps this could be an in-class activity with make-up for all students who miss the class.”

Warrant 1, that all targeted students complete BELA, was partially supported. Amongst the extant PELA literature, a completion rate of above 90% is noteworthy. However, as universal uptake was not achieved, the warrant was only partially supported.

5.1.5.2 WARRANT 2

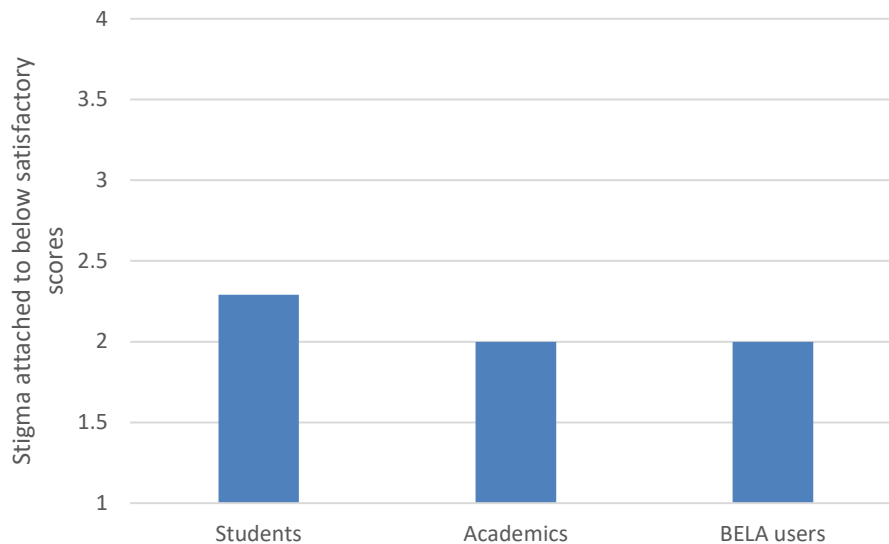
Warrant 2 requires the avoidance of any stigma and/or disadvantage for all students. This warrant involved analysis of stakeholder responses via the student questionnaire, academic staff questionnaire and BELA users questionnaire and feedback from external raters.

In order to assess whether BELA results in any stigma for students, responses to the item, “There is stigma attached to students whose writing is assessed as being below satisfactory” within the student questionnaire ($n = 99$), academic staff questionnaire ($n = 8$), and BELA users questionnaire ($n = 17$), were analysed. A one-way ANOVA was conducted

to compare the mean scores on stigma related with the task (DV) in relation to the three questionnaires (IV; students, academics, BELA users). There was no significant difference in scores on stigma related with the task between students ($M = 2.29$, $SD = 0.90$), academics ($M = 2.00$; $SD = 0.76$), and BELA users ($M = 2.00$; $SD = 0.71$), $F(2, 121) = 1.15$, $p = .322$. The mode for students, academics and BELA users was 2 (*Disagree*; $n_{Students} = 42$, 42.4%; $n_{Academics} = 4$, 50%; $n_{BELA\ users} = 12$, 70.6%). Although not definitive, these results imply that all groups of respondents, overall, believed there was no stigma associated with scoring below satisfactory on BELA (see Figure 5.30).

Figure 5.30

Mean opinion comparison of the stigma attached to poor performance for students, academics, and BELA users (N = 124)



Furthermore, external raters were asked if they predicted a stigma to be experienced by students who performed poorly on BELA. External rater 1 opined that, in theory, there should not be any stigma attached, “There shouldn't be, I mean, who would know anyway? Like nobody, nobody knows the results, do they? Well, just the students.” The rater elaborated and said, English-speaking background students who performed poorly may feel some stigma amongst their peers, stating:

“Certainly, the second language ones wouldn't [feel stigmatised]. The only stigma I could think of is with the native speakers. Like if, you know, if I'm mates with someone in the class, and then I'm told that, you know, I have to go and do extra writing help, even though I've just done 12 years of high school here, and my English

is my native language. And then maybe they would. There might be a stigma attached to them, perhaps. But I think the international students whose second language is English, I don't think they care. I think they'd be happy to get the extra attention.”

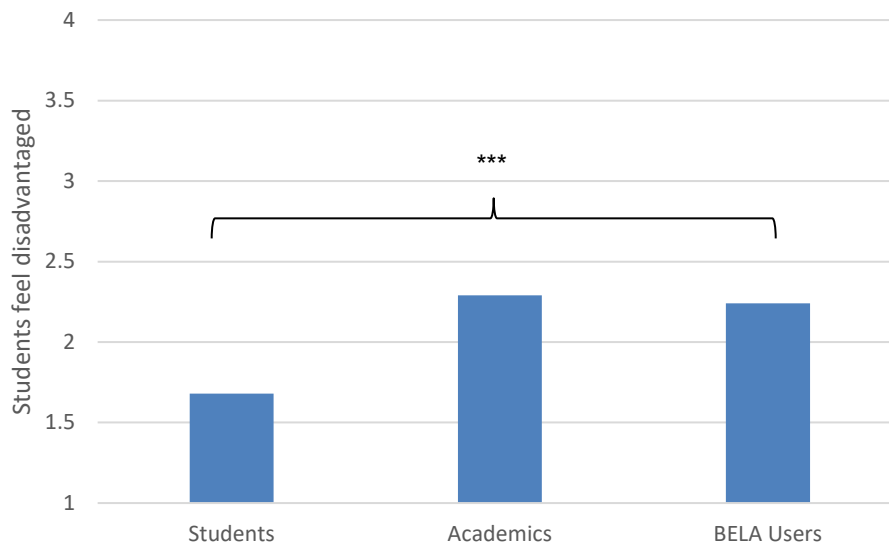
External rater 2 believed that there is a chance that some students may feel a “loss of face”. They explained:

“There may be with certain students, as I said, just based on personality generally or possibly family expectations, family pressure There could be students who have a slightly inflated view of their ability to write, or they may have just had a bad day that day and they just didn't, you know, write it as well as they could have Those students that we're talking about, they would be pretty rare. I think most students would see it as, it's been framed as a positive thing as something that the university is doing for you. Occasionally, you just get students, I think, who don't see it that way, and there can be a lot of reasons for that.”

To evaluate whether stakeholders believed BELA disadvantaged some students, a one-way ANOVA was run to compare mean scores on the belief that students felt disadvantaged (DV) in relation to the three different questionnaires (IV; students, academics, BELA users). Specifically, the items, “I felt disadvantaged when completing BELA” for students ($n = 99$) and “Some students feel disadvantaged when completing BELA” for academic staff ($n = 7$) and BELA users ($n = 17$) were included. Here, there was a significant difference between students ($M = 1.68, SD = 0.73$), academic staff ($M = 2.29; SD = 0.95$), and BELA users ($M = 2.24; SD = 0.44$), $F(2, 120) = 6.35, p = .002$. The mode for students was 1 (*Strongly disagree*; $n = 45, 45.5\%$); in contrast, the mode for academics was 3 (*Agree*; $n = 4, 57.1\%$), and the mode for BELA users was 2 (*Disagree*, $n = 13, 76.5\%$). This can be considered a medium effect size ($f = .31$; Cohen, 1988). After post-hoc analysis, there was a significant difference between students and BELA users. In this case, it indicated that students strongly disagreed that some felt disadvantaged when completing BELA, but academic staff and BELA were uncertain (see Figure 5.31).

Figure 5.31

Mean opinion comparison of whether students feel disadvantaged when completing BELA for students, academics, and BELA users (N = 124)



Note. *** $p < .001$.

Warrant 2, that BELA does not cause any stigma or disadvantage for students, was supported. External raters opined that some may have been stigmatised or embarrassed by poor performance on BELA, but this was not reflected in the student data. Students, most importantly, disagreed that there was a stigma attached to performing poorly on BELA and indicated they had not experienced disadvantage.

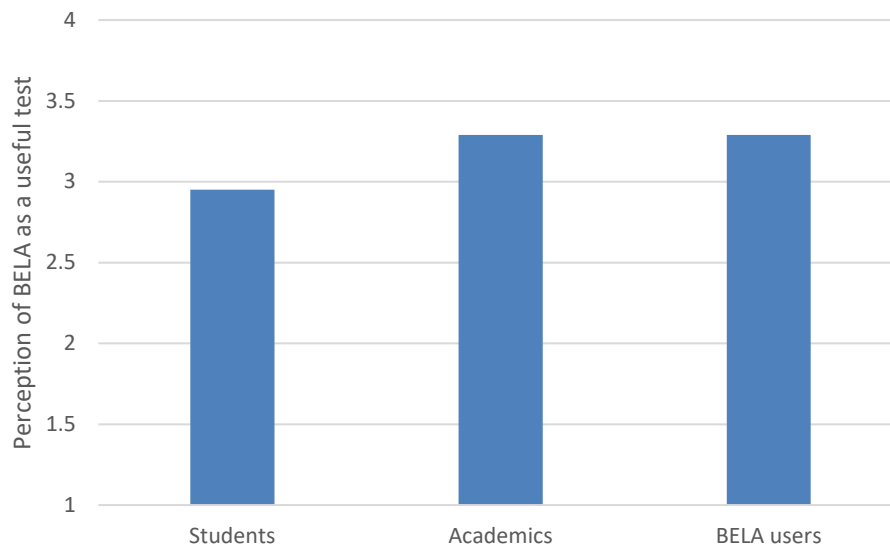
5.1.5.3 WARRANT 3

Warrant 3 states that students should perceive BELA and its usefulness in a positive manner. In terms of student feedback, the student questionnaire and the ASC feedback surveys were analysed. Furthermore, feedback from staff stakeholders (i.e., academic staff and BELA users) via relevant questionnaires was compared to student responses.

To assess whether test takers believed that BELA was a useful task, the answers to the question “The usefulness of BELA is clear” (DV) from the IVs, the student questionnaire ($n = 99$), academics questionnaire ($n = 7$) and the BELA users questionnaire ($n = 17$) were compared using a one-way ANOVA. There was a non-significant difference in mean scores between students ($M = 2.95$, $SD = 0.66$), academics ($M = 3.29$; $SD = 0.49$), and BELA users ($M = 3.29$; $SD = 0.59$), $F(2, 120) = 2.73$, $p = .069$, implying that all three stakeholders, overall, perceived BELA as a useful task (see Figure 5.32).

Figure 5.32

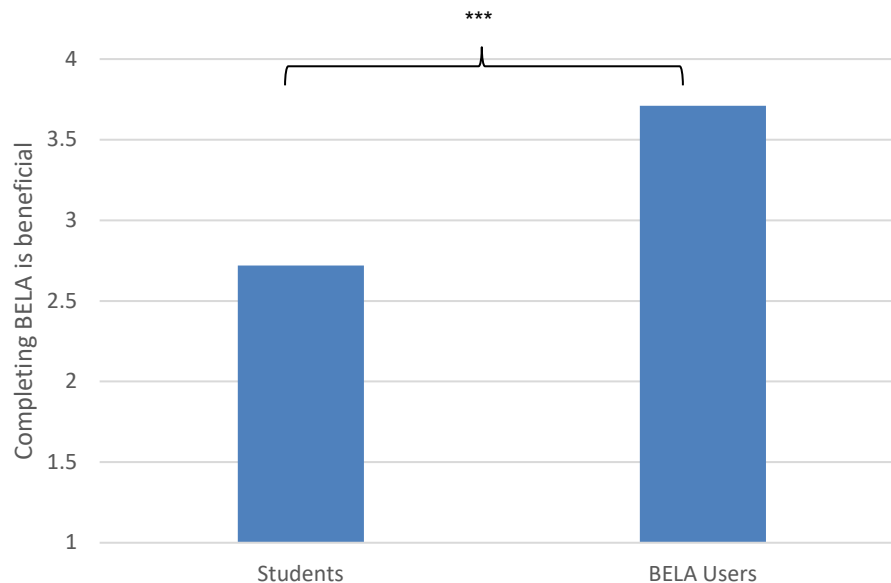
Mean comparison of perception of BELA as a useful test between students, academics, and BELA users (N = 123)



In addition, to compare whether students and BELA users believed completing the PELA was useful, responses to the items, “Completing BELA has been beneficial for me” for students ($n = 99$) and “Completing BELA is beneficial for students” for BELA users ($n = 17$) were analysed using a two-tailed, independent samples t-test. There was a significant difference in mean scores between students ($M = 2.72$, $SD = 0.80$) and BELA users ($M = 3.71$; $SD = 0.47$), $t(109) = -4.96$, $p < .001$. This can be considered a large effect size ($d = 1.30$; Cohen, 1988). The mode for students was 3 ($n = 50$, 50.5%), while the mode for BELA users was 4 ($n = 12$, 70.6%), indicating that students and BELA users disagreed. On one hand, students were uncertain whether BELA was beneficial for them, whilst BELA users strongly agreed that completing BELA benefitted students (see Figure 5.33).

Figure 5.33

Mean comparison of perception of BELA as beneficial between students and BELA users (N = 116)



Note. *** $p < .001$.

Additionally, to determine whether BELA users ($n = 17$) considered BELA a useful measure, the responses to the item, “BELA is a useful measure of students' academic writing ability” were analysed. BELA users agreed with the statement ($M = 3.41$, $SD = .51$), indicating that BELA was seen as a useful measure of students' academic writing amongst the stakeholder group.

Qualitative feedback regarding the usefulness of the PELA was also obtained via the students questionnaire. Overall, feedback was positive. The most common positive aspect of BELA students referred to was the ability for students themselves to use BELA as an instrument to “check” their academic writing skills. One student maintained, “I thought this was good as i didn't know what standard my writing was at before i completed it (sic). The tips were also really helpful.” Another student indicated, “Was good to be tested to make sure I can write well and that I will be supported.” Finally, a student expressed that BELA “showed me how I was sitting coming into university.”

Additional qualitative feedback was obtained from students via the open-ended item in the ASC feedback survey. Most students who responded to the feedback survey perceived BELA and their experience completing it positively. Notable comments from students were

that BELA and the associated feedback increased students' academic confidence. One student explained, "It was helpful and gave me confidence in my writing ability", while another student expressed, "it was Good (sic) to reinforce my writing skills." Students also noted the value in completing BELA as a transition from previous studies (e.g., high school) to university. One student maintained, "I found it helpful in understanding what type of academic writing was expected for university. It was a nice transition." Additional indicative quotes concerning the transition to university studies included, "nice way to start first uni assignment" and "Good activity to her [get] back into it."

There were, however, students who expressed negative attitudes towards BELA via the ASC feedback survey. Two students indicated that the assessment was not necessary (i.e., "I don't think that it is necessary at all" and "It was unnecessary"). Moreover, one student indicated that BELA had not been beneficial for them. The student expressed, "It was a good introduction to the CORE class, but it hasn't benefitted me in any way." Another also commented that it was not useful for them as they were enrolled in a degree that did not require substantial amounts of academic writing (e.g., Bachelor of Architecture). However, the student acknowledged that completing BELA was potentially useful for other students. The student explained, "I can understand that it could be useful for international students, however, I do not think it was particularly useful for me, especially as my degree doesn't require much writing." The belief that BELA was beneficial for others, even if it did not directly benefit themselves, appeared to be a common theme. Another student explained, "The idea is good. I believe it would be beneficial for people who severely lack in an area of academic writing. However, for the majority of university students it was a simple quick task done before class. I do feel it's (sic) presence is important for the small percentage who need help however."

To further understand students' beliefs about the benefit of completing BELA, student responses ($n = 99$) to the item, "BELA benefits other students", were analysed. Students agreed with the statement concerning other students ($M = 3.04$, $SD = 0.66$) with a mode of 3 (*Agree*; $n = 70$, 70.7%). This suggests that students agreed that BELA was beneficial for others.

Qualitative feedback was finally obtained via the BELA users questionnaire. The feedback received from BELA users was very positive. One BELA user expressed, "it is

seems (sic) a useful tool for identifying students at academic risk.” Others spoke more highly of the assessment and its associated processes, for instance, arguing:

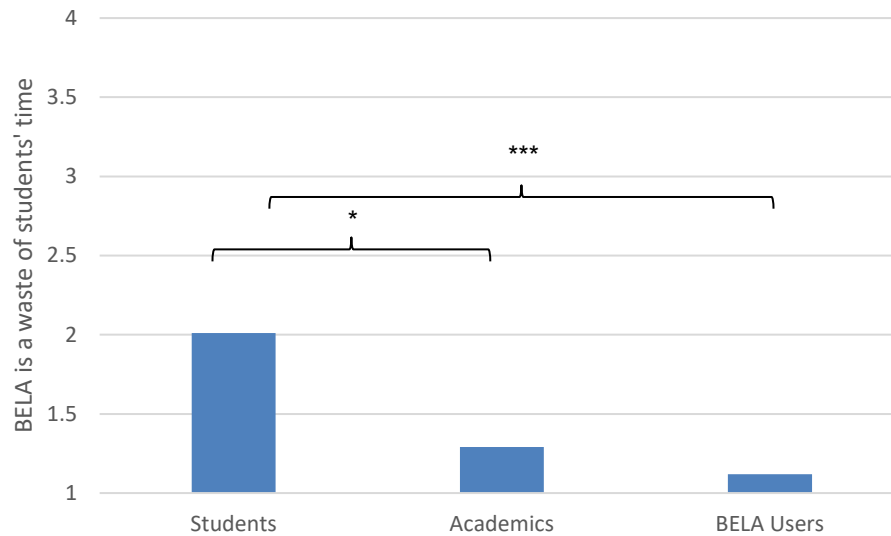
“I think BELA is a great tool to use at the beginning of both UG [undergraduate] and PG [postgraduate] degrees. It is an easy task for students whose writing is above satisfactory but even for them the test sets a standard so they understand the expectations. It helps students who need help to identify the needs early and it encourage them to meet with SLS [now Academic Skills Centre] which is great help especially for students who need the extra help to pass/improve their assessments.”

However, it was noted that BELA could not measure everything. One BELA user argued, “I think the BELA is extremely useful and gives an indication to a students (sic) transitional needs. However, the one variable that it cannot measure is a students (sic) motivation to complete the task and their perception of the task.”

Furthermore, answers to the item “BELA was a waste of my time” (DV) from the IVs, the student questionnaire ($n = 99$) and “BELA is a waste of students' time” from both the academic questionnaire ($n = 7$) and BELA users questionnaire ($n = 17$) were compared using a one-way ANOVA. There was a significant difference in mean scores between students ($M = 2.01$, $SD = 0.78$), academics ($M = 1.29$; $SD = 0.49$), and BELA users ($M = 1.12$; $SD = 0.33$), $F(2, 120) = 13.39$, $p < .001$. This can be considered a large effect size ($f = .59$; Cohen, 1988). Post hoc analyses with Tukey's HSD (using an α of .05) revealed a significant difference in opinion between students and academics ($p = .03$), and students and BELA users ($p < .001$), implying that each stakeholder group clearly disagreed that BELA was a waste of students' time. However, disagreement was far stronger for staff stakeholders, in comparison to students (see Figure 5.34).

Figure 5.34

Mean comparison of perception of BELA as a waste of students' time between students, academics, and BELA users (N = 123)



Note. * $p < .05$, *** $p < .001$.

Warrant 3, that students' perceptions of BELA and its usefulness are positive, was supported. Although perceived benefits of BELA may have been lacking amongst some students, the majority of feedback from students indicated that BELA was useful. Staff stakeholders strongly agreed that BELA was a useful and purposeful assessment.

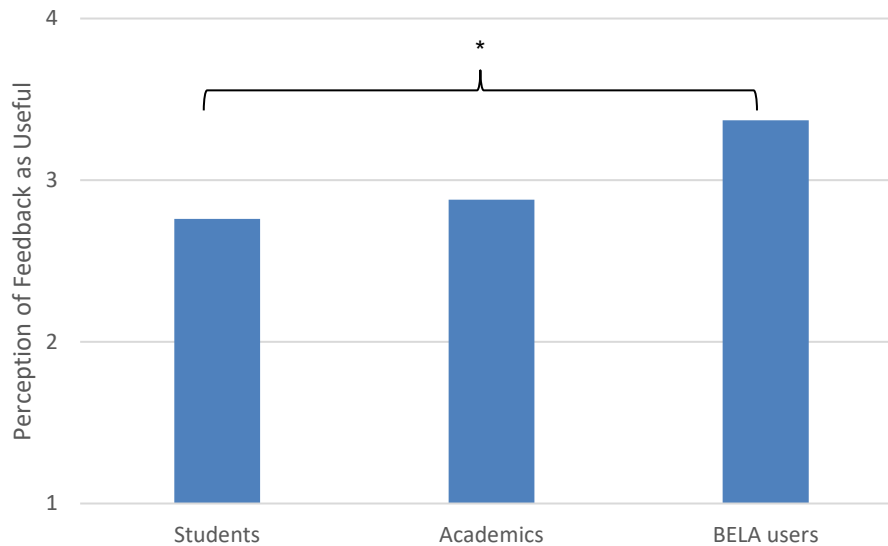
5.1.5.4 WARRANT 4

Warrant 4 focuses on the usefulness of feedback from BELA and necessitates that it informs future learning. This firstly involved analysing responses to the item "The feedback I received after completing my BELA has helped me to develop my writing" (DV) for students and the item, "The feedback students receive after completing BELA helps them develop their writing" for academics and BELA users (IVs) using a one-way ANOVA. There was a significant difference in mean scores between students ($M = 2.76$, $SD = 0.81$), academics ($M = 2.88$; $SD = 0.84$), and BELA users ($M = 3.37$; $SD = 0.50$), $F(2, 123) = 2.98$, $p = .008$. The mode for students and BELA users was 3 (*Agree*; $n_{Students} = 50$, 50.5%; $n_{BELA\ users} = 12$, 63.2%), whilst for academic staff, multiple modes existed, 2 (*Disagree*; $n = 3$, 37.5%) and 3 (*Agree*; $n = 3$, 37.5%). This can be considered a medium effect size ($f = .30$; Cohen, 1988). Post hoc analyses with Tukey's HSD revealed a significant difference in opinion between students and BELA users, indicating uncertainty amongst students and academic staff. In

contrast, BELA users believed that BELA was able to inform students' learning (see Figure 5.35).

Figure 5.35

Mean comparison of perception of feedback as useful and informing future learning between students, academics and BELA users (N = 126)



Note. * $p < .05$

In addition to the above analysis, academics were asked whether students' BELA essays informed their teaching. Specifically, the responses of academics ($n = 7$) to the item, "Students' writing in BELA can inform instruction", were analysed. Most academics agreed with the statement ($M = 3.00$, $SD = .58$), indicating that the students' writing produced via BELA assisted their teaching. A similar question was posed to BELA users ($n = 17$). The responses to the item, "Feedback from BELA directly informs students' future learning", indicated that BELA users agreed feedback from BELA informed student learning ($M = 3.12$; $SD = 0.70$).

Feedback was also gathered from students' qualitative responses to the questionnaire. A raised awareness of academic essay writing conventions via the BELA process was noted by multiple students via the student questionnaire. One student stated that BELA was "Very useful. I am grateful for the BELA because it made me aware of how to structure Academic Writing." Another student expressed, "Completing and reviewing the BELA helped to remove any unnecessary elements that I had previously attached with my style of writing.

The BELA is a great initial writing task to ensure academic style is established early in university life.” This evidence is supportive of the statement that BELA and follow-up feedback informed students’ future learning.

Additional qualitative responses from the ASC feedback survey were analysed. Students expressed that they received valuable feedback from completing BELA and that the feedback informed future learning. One student stated, “it was a good introduction into University basic expectations.” Another student commented, “It was a good way to gage (sic) how my writing skills were and if I needed to put more effort into improving them.” Finally, a student noted the value in completing BELA, stating, “The BELA assignment is a great way to assess current writing skills and abilities. I found the experience helpful and the feedback valuable.” However, as noted regarding Warrant 2 of the Decision inference, multiple students had expressed that there was a lack of feedback; for instance, “It would have been useful to know what I should have improved on.” Thus, further investigation regarding students’ perception of feedback is necessary.

The external raters were asked during semi-structured interviews whether BELA could inform future teaching and learning. Initially, External rater 1 was somewhat undecided, stating:

“Well, I'd hope the students can improve. I mean, I think it's, you know, I don't know what the lecturers’ attitude towards that is. But I think, I mean really, I just think it's really a useful tool basically to identify students whose writing needs some extra assistance, that's obvious. As I say, I'm not sure if it would inform teaching or if it should or not. That's just a very useful tool.”

In contrast, however, External rater 2 clearly indicated that BELA can inform future teaching and learning, and that the measure was a valuable tool. They argued:

“Yeah, I think actually insights that you gained from BELA are not just useful for Learning Advisors [at Academic Skills Centre] who maybe are tasked with assisting students who perhaps don't perform so well, but I think insights you gained from the BELA are really useful for the broader academic community at Bond. So firstly at Bond, but also beyond that. So, I think it's really useful for, you know for example, the English language institute. It's important that they know what the performance of these students is like. So, there are certain things. Perhaps they may need to change in

their offerings, in their teaching. It's really great for academics to see what students can produce because sometimes they have a slightly unrealistic expectation of what students can do. Sometimes they, I think, think students can do more, but other times, I think they actually underestimate students. So, I think, yeah, I think it's really useful to disseminate that information amongst colleagues and then and then beyond the university as well. I think PELAs are a really great thing. I think all universities should have them in Australia. And so I think, the insights that you gain from this project hopefully will encourage other institutions to introduce them and use them.”

Warrant 4, that feedback from BELA is useful and directly informs future learning, was partially supported. Responses to questionnaire items indicated partial agreement amongst students. Qualitative responses from students demonstrated BELA had informed their understanding of essay writing conventions through “valuable” feedback. Academic staff and BELA users demonstrated their belief that BELA informed teaching. However, as several students indicated there was a lack of feedback, this warrant was partially supported.

5.1.5.5 WARRANT 5

Warrant 5 requires students to act on recommendations as a consequence of their BELA result. In response, uptake rates were analysed via internal data of students utilising the ALL unit, ASC. Of the 190 students who were flagged as having Below satisfactory academic essay writing, 87.34% ($n = 166$) attended ASC. Similar to BELA completion rates, although a high percentage of students who may need assistance with academic essay writing utilised the on-campus support service, universal uptake of recommended support was not achieved. Hence, this warrant was only partially supported.

5.1.5.6 WARRANT 6

Warrant 6 concerns the appropriacy of follow-up academic language development options. Analysis of student feedback from the student questionnaire and the ASC feedback surveys, in comparison to academic staff and BELA user responses, was required.

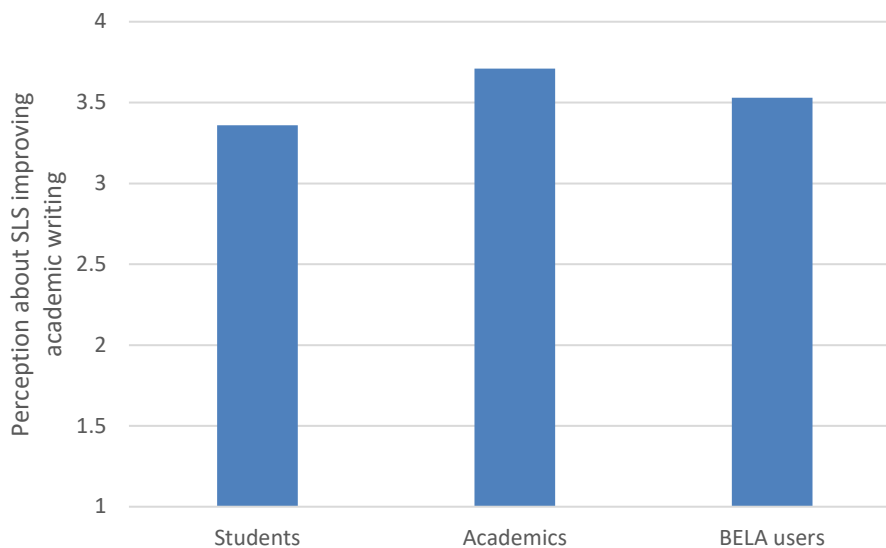
First, to ascertain whether the follow-up language development options provided for students were appropriate, the answers to the question “As a consequence of attending Student Learning Support, my academic writing has improved” (DV) of students, academics and BELA users (IVs) were compared using a one-way ANOVA. There was a non-significant difference in mean scores between students ($M = 3.36$, $SD = 0.76$), academic staff

($M = 3.71$; $SD = 0.49$), and BELA users ($M = 3.53$; $SD = 0.51$), $F(2, 57) = 0.97$, $p = .387$.

This demonstrated that all three groups believed that attending ASC (i.e., Student Learning Support) helped improve students' academic writing (see Figure 5.36).

Figure 5.36

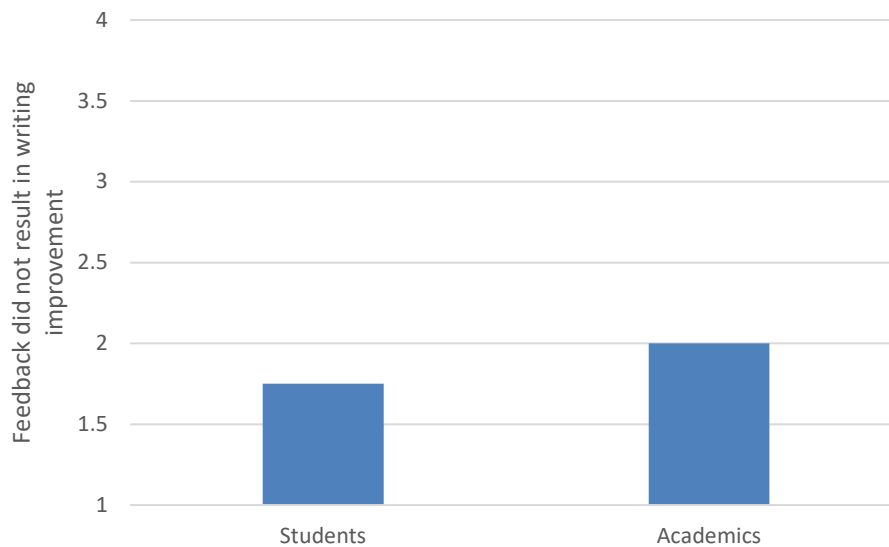
Mean comparison of perception of SLS (ASC) to improve academic writing between students, academics and BELA users ($N = 60$)



Second, to confirm stakeholders' viewpoints regarding academic writing development, responses to the items, "Even though I received feedback and support, my academic writing has not improved" for students ($n = 36$) and "Even though students receive feedback and support, their academic writing does not improve" for academic staff ($n = 7$) were analysed using a two tailed, independent samples t-test. There was a non-significant difference in mean scores between students ($M = 1.75$, $SD = 0.55$) and academics ($M = 2.00$; $SD = 0.58$), $t(41) = -1.09$, $p = .284$, indicating that both students and academics disagreed that receiving feedback and support had not assisted in developing students' academic writing (see Figure 5.37).

Figure 5.37

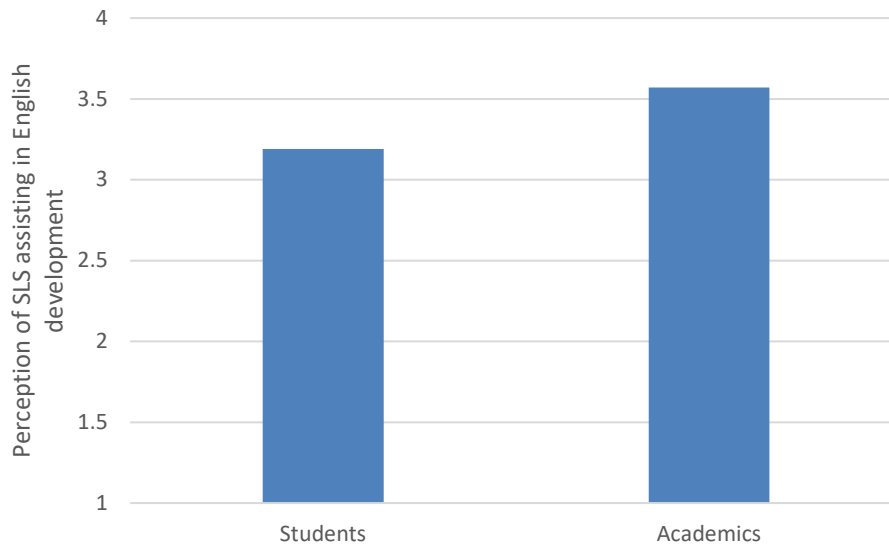
Mean comparison of perception that although feedback and support was provided, academic writing did not improve ($N = 43$)



Third, answers to the items, “My English has improved as a result of attending Student Learning Support” for students ($n = 36$) and “Students’ English improves as a result of attending Student Learning Support” for academics ($n = 7$) were also compared using a two tailed, independent sample t-test. There was a non-significant difference in mean scores between students ($M = 3.19$, $SD = 0.89$) and academics ($M = 3.57$; $SD = 0.54$), $t(41) = -1.08$, $p = .287$, implying that both students and academics believed that attending ASC (i.e., Student Learning Support) helped improve students’ English (see Figure 5.38).

Figure 5.38

Mean comparison of perception of SLS (ASC) and English improvement between students and academics (N = 43).



Regarding students' qualitative feedback, the majority of responses to the open-ended items in the questionnaire were highly positive of support provided to them. Indicative quotes included, "I found the help received from the Academic Skills Centre invaluable", "... I attended learning support prior to this and also an academic writing workshop so that I could improve on my writing skills immediately. I believe this definitely helped with receiving a better mark on my BELA" and "Thanks SLS for providing such great support!"

Further positive responses were found in the ASC feedback survey. One student explained, "I did the BELA writing task before my first sls session and discussed it with [Learning Advisor's name]. he gave me useful tips concerning academic writing." Another student stated, "Im not going to lie i (sic) did poorly in the BELA and was referred here for further help and since then I never wanted to stop coming to SLS its been a huge help." Finally, another student indicated, "It was great to practice writing an essay at a university level as i was fresh from high school (sic), [Learning Advisor's name] was amazing at helping me be more concise and reference properly."

Staff stakeholders (i.e., academic staff and BELA users) agreed with students that ASC provided valuable support. One academic noted, "All the advisers at SLS are doing great job (sic) in helping our students in many different areas. Students who actively seek

help almost always get higher marks and definitely always learn more and better.” Meanwhile, a BELA user expressed that BELA “... helps students who need help to identify the needs early and it encourage them to meet with SLS which is great help especially for students who need the extra help to pass/improve their assessments.” A final BELA user noted the important role BELA played and argued:

“It is important that we continue to have BELA services. We are facing a generation of students where English language is an option in their high school curricula. While they may be technically proficient in responding to specific field of studies, it pains me to come across emails with poor sentence structure, not putting punctuation in the right places or not bothering to make the proper address which I thought are fundamentals to writing. BELA process works because of the efficiency.”

Overall, Warrant 6, that follow-up writing development options are appropriate, was supported. All stakeholders expressed that students’ academic writing and language skills improved with the support of ASC consultations.

5.1.5.7 WARRANT 7

Warrant 7 requires the development of academic essay writing skills amongst students who utilised language support. To ascertain whether development occurred and if support from ASC played a role in such development, a matched sample analysis was conducted. This involved assessing whether gains occurred (i.e., comparing BELA and BELA 2 results) for students who had utilised the support service during the semester (i.e., attending at least one follow-up consultation) in comparison to students who had not. In total, 82 students completed the non-compulsory BELA 2. Of these 82 students, five had not completed the original BELA in Core 1: Critical Thinking and Communication; thus, these five were removed from the data analysis, leaving 77 students’ BELA data for analysis. BELA 2 results, as well as number of students and percentage of students achieving each score, are displayed in Table 5.9.

Table 5.9*Breakdown of scores on BELA 2*

BELA 2 result	Number of students	Percentage
3	2	2.6%
5	14	18.2%
7	29	37.7%
9	32	41.6%

Regarding attendance at ASC, 37 students had completed the original BELA in Core 1: Critical Thinking and Communication, BELA 2 and attended at least one consultation ($M = 3.37$, $SD = 3.56$) at ASC. In comparing these students' results on BELA and BELA 2, eight out of 10 students in the Below satisfactory group (scores of 3s and 5s) who utilised ASC demonstrated improvement in their academic essay writing via BELA 2. Additionally, seven of the 13 students scoring 7 on the original BELA, improved their writing in BELA 2. This demonstrates that, overall, most students who performed poorly on BELA but attended at least one ASC appointment developed their academic essay writing over the course of the semester, as demonstrated by BELA 2.

In comparison, the students who had completed both the original BELA and then BELA 2 but had not engaged with ASC ($n = 40$), were used as a matched sample to the group who had engaged ASC. In comparing these students' BELAs, four out of six students whose writing was considered to be Below satisfactory category improved in BELA 2, while five students in the Satisfactory category out of 15 improved in BELA 2 (see Table 5.10). This demonstrates that students in this group also developed their academic essay writing over the course of the semester. However, more students in the group that had attended ASC demonstrated gains in academic writing compared to the group that had not engaged ASC. These results can be seen in Table 5.10 below; please note, as BELA 9 is the highest score achievable, students scoring 9 on the original BELA were excluded, as improvement with this assessment was not possible.

Table 5.10*Proportion of students in each BELA category who improved in BELA 2*

BELA 1 score	Number of students who improved by at least 1 band in their BELA 2 result.			
	Attended 1 or more consultations at ASC	%	Did not attend consultations at ASC	%
3	1 out of 1	100%	0 out of 1	0%
5	7 ^a out of 9	77.8%	4 out of 5 ^b	80%
7	7 out of 13	53.8%	5 out of 16	31.3%

^a Two of these students improved by two bands (i.e., achieving 9 on BELA 2).

^b One of these students improved by two bands (i.e., achieving 9 on BELA 2).

Warrant 7, that students who take up support options improve their English, was supported. Most students who completed a second BELA developed their academic writing over the course of one semester at university. This was particularly the case for students who attended at least one ASC appointment.

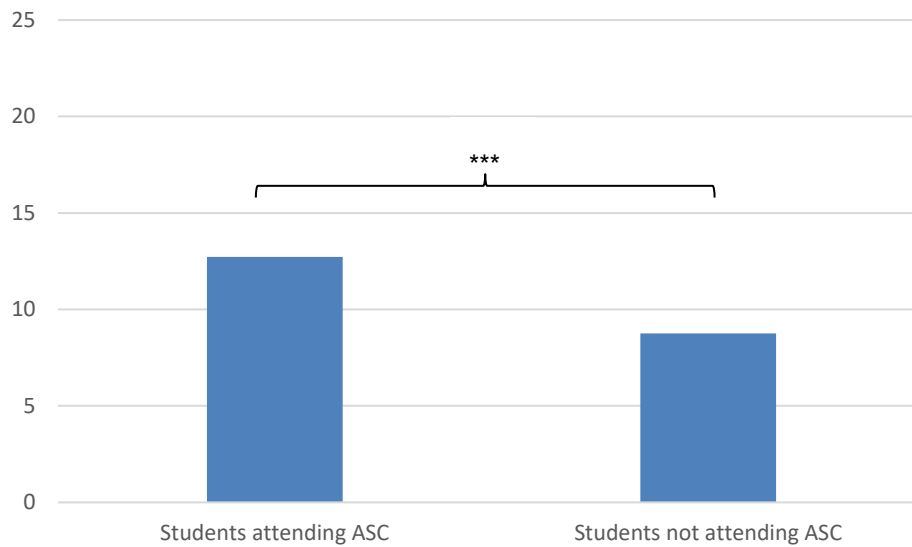
5.1.5.8 WARRANT 8

Warrant 8 suggests that students who do not act on the recommendations regarding language development may have inferior academic outcomes compared to students who do. Examining this warrant required analysing attendance data from ASC and academic performance of students in the Below satisfactory BELA category who attended ASC consultations in comparison to those who had not. Specifically, academic performance between the two groups was compared for: 1) the Major Essay in Core 1: Critical Thinking and Communication, 2) Overall Grade in the subject, and 3) GPA after two semesters. Additionally, to determine a relationship between attendance at ASC and academic performance, Point Biserial correlations were run.

First, to compare performance on the Major Essay amongst students categorised as having Below satisfactory BELA scores (i.e., scores of 3 and 5) and who attended at least one ASC consultation ($n = 160$) in comparison with students with Below satisfactory BELA scores and who did not attend ASC consultations ($n = 31$), a two tailed, independent samples t-test was run. There was a significant difference in Major Essay marks with students who attended ASC scoring higher on the Major Essay ($M = 12.72$, $SD = 5.86$) in comparison to those who did not attend ASC ($M = 8.76$, $SD = 6.50$), $t(189) = 3.39$, $p < .001$ (see Figure 5.39). This can be considered a medium effect size ($d = .66$; Cohen, 1988).

Figure 5.39

Major Essay marks amongst students categorised as having Below satisfactory BELA scores (N = 191)



Note. *** $p < .001$.

Figures 5.40 to 5.41 illustrate the distribution of Major Essay grades for students categorized as having Below satisfactory writing and 1) attended ASC and 2) not attending ASC. Notably, students receiving Below satisfactory BELA scores and attending ASC were negatively skewed (-.863), whilst students obtaining Below satisfactory BELA scores and not attending ASC were more less negatively skewed (-.471).

Figure 5.40

Histogram illustrating distribution of Major Essay results in Below satisfactory category who attended ASC (N = 160)

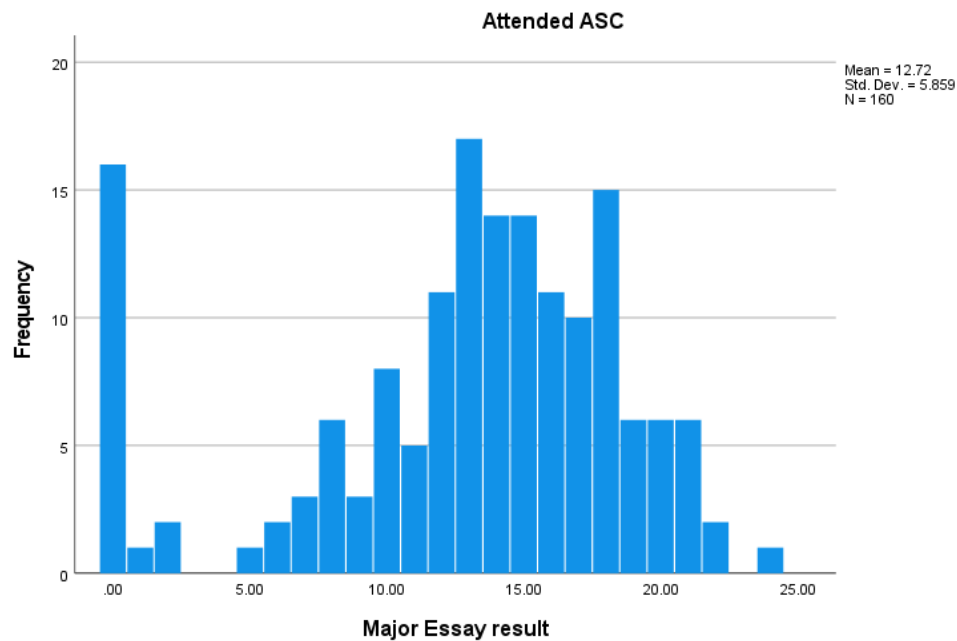
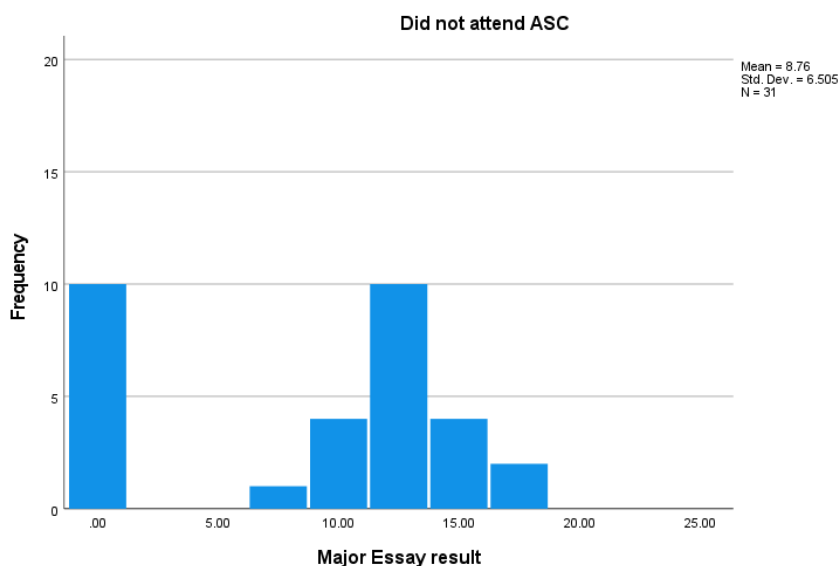


Figure 5.41

Histogram illustrating distribution of Major Essay results in Below satisfactory category who did not attend ASC (N = 31)

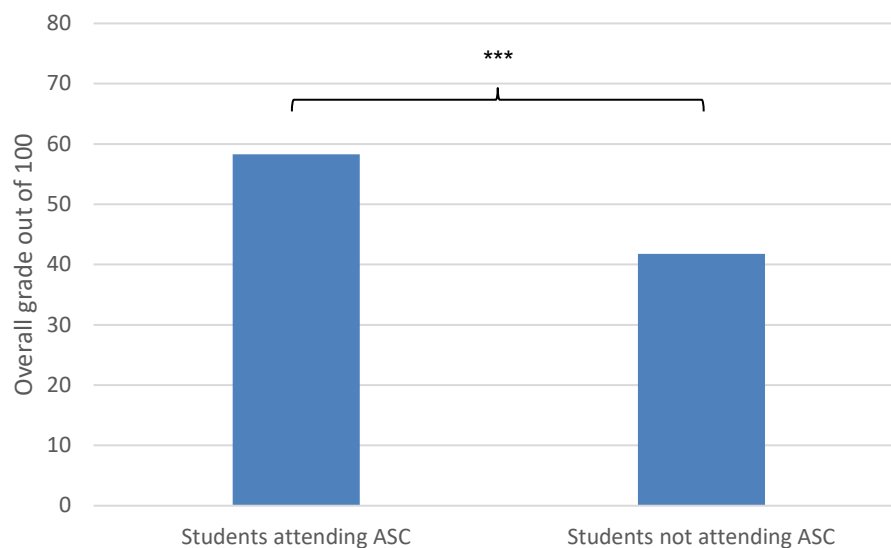


Second, to compare performance on overall in the subject, Core 1: Critical Thinking and Communication, amongst students categorised as having Below satisfactory BELA scores (i.e., scores of 3 and 5) and who attended at least one ASC consultation ($n = 160$) in

comparison with students with Below satisfactory BELA scores and who did not attend ASC consultations ($n = 31$), a two tailed, independent samples t-test was run. There was a significant difference in overall grades with students who attended ASC scoring higher overall ($M = 58.30, SD = 17.02$) in comparison to those who did not attend ASC ($M = 41.76, SD = 18.88$), $t(189) = 4.86, p < .001$ (see Figure 5.42). This can be considered a large effect size ($d = .95$; Cohen, 1988).

Figure 5.42

Overall grade amongst students categorised as having Below satisfactory BELA scores ($N = 191$)



Note. *** $p < .001$.

Figures 5.43 to 5.44 below illustrate the distribution of Overall Grades for students categorized as having Below satisfactory writing and 1) attended ASC and 2) not attending ASC. Notably, students' Overall Grades for those receiving Below satisfactory BELA scores and attending ASC were negatively skewed ($-.909$), whilst students obtaining Below satisfactory BELA scores and not attending ASC were less negatively skewed ($-.632$).

Figure 5.43

Histogram illustrating distribution of Overall Grades in Below satisfactory category who attended ASC (N = 160)

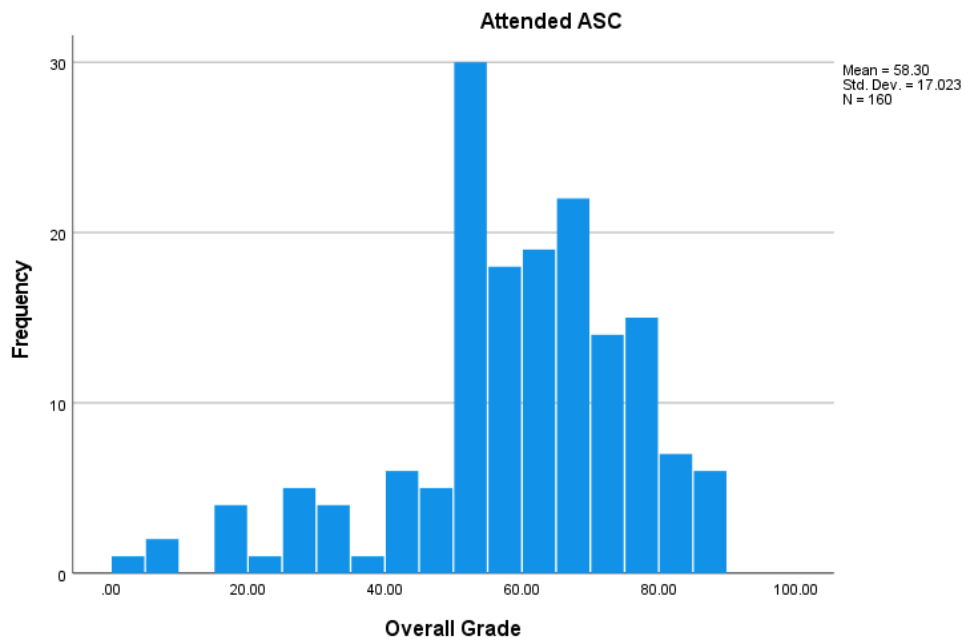
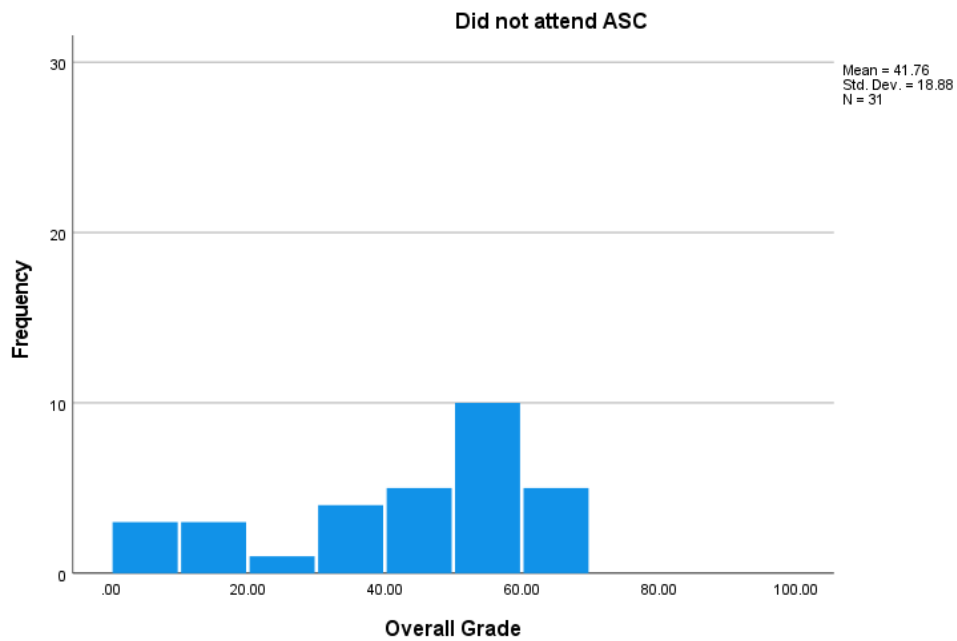


Figure 5.44

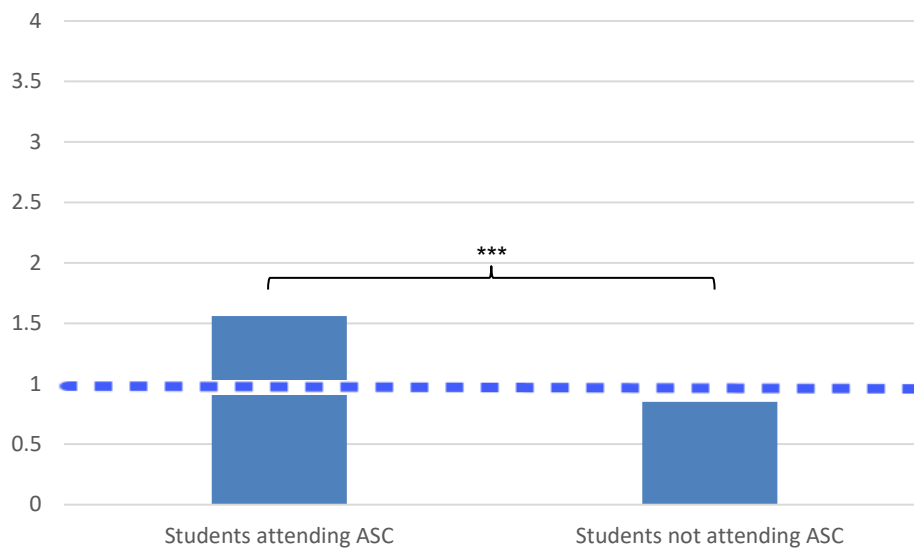
Histogram illustrating distribution of Overall Grades in Below satisfactory category who did not attend ASC (N = 31)



Third, to compare GPA after two semesters amongst students categorised as having Below satisfactory BELA scores (i.e., scores of 3 and 5) and who attended at least one ASC consultation ($n = 160$) in comparison with students with Below satisfactory BELA scores and who did not attend ASC consultations ($n = 31$), a two tailed, independent samples t-test was run. There was a significant difference in GPA with students who attended ASC maintaining a higher GPA ($M = 1.56$, $SD = 0.86$) in comparison to those who did not attend ASC ($M = 0.85$, $SD = 0.66$), $t(189) = 4.33$, $p < .001$ (see Figure 5.45; please note the dotted line represents GPA of 1.0 or a pass grade). This can be considered a large effect size ($d = .85$; Cohen, 1988).

Figure 5.45

GPA after two semesters amongst students categorised as having Below satisfactory BELA scores ($N = 191$)



Note. *** $p < .001$.

Figures 5.46 and 5.47 below illustrate the distribution of GPA for students categorised as having Below satisfactory writing and 1) attended ASC and 2) not attending ASC. GPA of 1.0 (Pass) is indicated by the dotted line. Notably, students' GPA for those receiving Below satisfactory BELA scores and attending ASC were negatively skewed (-.405), whilst students obtaining Below satisfactory BELA scores and not attending ASC were positively skewed (.266).

Figure 5.46

Histogram illustrating distribution of GPA in Below satisfactory category who attended ASC (N = 160)

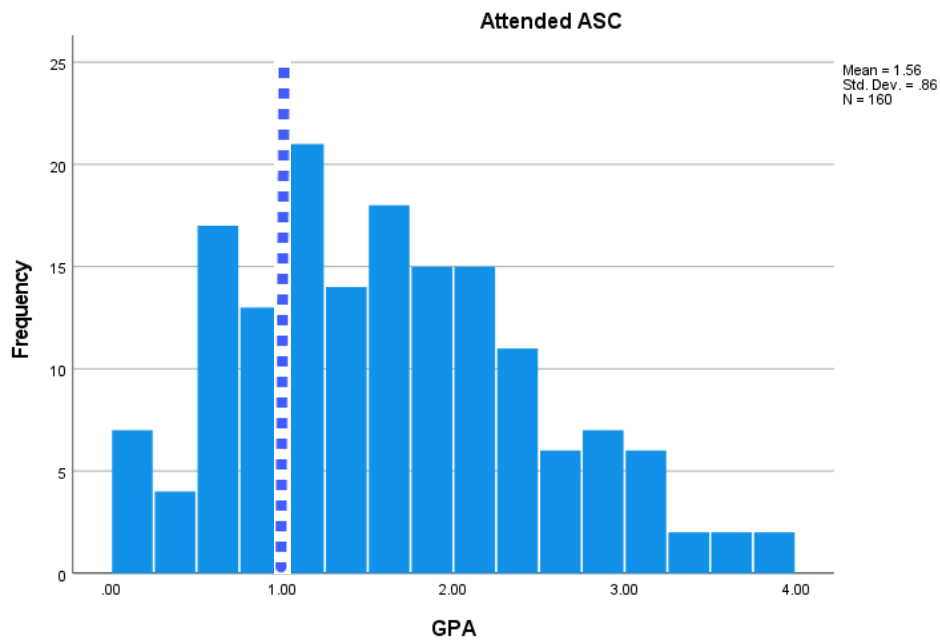
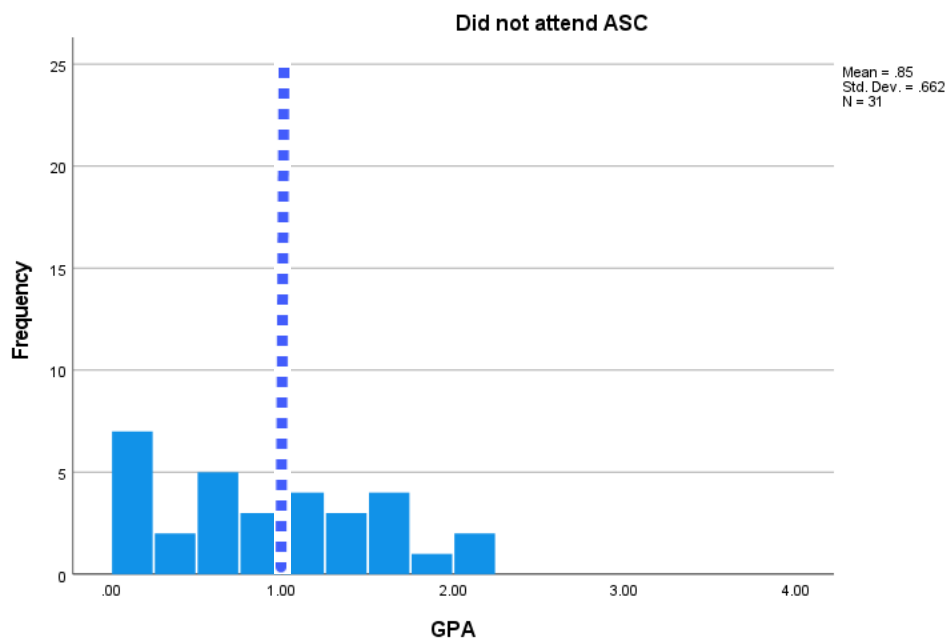


Figure 5.47

Histogram illustrating distribution of GPA in Below satisfactory category who did not attend ASC (N = 31)



Finally, to assess whether students who failed to act on test recommendations (i.e., attend ASC) were more likely to struggle in their academic studies, a series of Point Biserial correlations was run. Students who did not attend ASC (1 = Did not attend; 0 = Did attend) significantly scored lower in their Major Essay ($r_{pb} (249) = -.42, p < .001$), in their Overall Semester Grade ($r_{pb} (249) = -.55, p < .001$), and had significantly lower GPA after two semesters ($r_{pb} (249) = -.47, p < .001$). Not attending ASC, for students with Below satisfactory essay writing skills, explained 17.6%, 30.2%, and 22.1% of variance for the Major Essay, Overall Semester Grade and GPA after two semesters.

Warrant 8, that students who did not act on recommendations to attend ASC would be more likely to struggle academically, was supported. In each academic performance indicator (i.e., Major Essay, Overall Grade and GPA after two semesters), students categorised as having Below satisfactory academic essay writing demonstrated lower results, on average, compared to their peers who were also categorised as having Below satisfactory writing but who had attended at least one ASC consultation.

5.1.5.9 SUMMARY OF RQ1E

This section analysed the evidence gathered to respond to RQ1e, whether the consequences of using BELA and the decisions informed are beneficial to all stakeholders, aligned with the Consequences inference. Table 5.11 summarises the warrants and evidence sought to respond to research question RQ1e.

Table 5.11

Warrants and evidence for the Consequences inference for BELA

Consequences inference		
RQ1e: Are the consequences of using BELA and the decisions informed by BELA beneficial to all stakeholders?		
Warrants	Evidence	Supported?
1. All target test takers sit the test.	Internal data from Core 1 subject site (i.e., completion rates)	Partially supported
2. The test does not result in any stigma or disadvantage for students.	Student questionnaire responses	Yes

Consequences inference

RQ1e: Are the consequences of using BELA and the decisions informed by BELA beneficial to all stakeholders?

	Academic questionnaire responses BELA user questionnaire responses External rater feedback ASC feedback survey responses	
3. Test takers' perceptions of the test and its usefulness are positive.	Student questionnaire responses Academic questionnaire responses BELA user questionnaire responses ASC feedback survey responses	Yes
4. The feedback from the test is useful and directly informs future learning.	Student questionnaire responses Academic questionnaire responses BELA user questionnaire responses ASC feedback survey responses External rater feedback	Partially supported
5. Students act on the test recommendation.	Internal data from ASC (i.e., uptake rates)	Partially supported
6. Follow-up language development options provided for students is appropriate.	Student questionnaire responses BELA user questionnaire responses	Yes

Consequences inference

RQ1e: Are the consequences of using BELA and the decisions informed by BELA beneficial to all stakeholders?

	ASC feedback survey responses	
7. Learners taking up support options improve their English over the course of their studies.	Matched sample analysis to determine gains on BELA 2	Yes
8. Students who fail to act on test recommendations are more likely to struggle in their academic studies.	Correlational analysis using internal data from ASC (i.e., uptake rates) and academic performance data	Yes

As shown in the table, support was evident for warrants 2, 3, 6, 7, and 8. Partial support was found for Warrant 1, “All target test takers sit the test”, as approximately 95% of target test takers completed BELA. Similarly, Warrant 4, “The feedback from the test is useful and directly informs future learning”, was partially supported. Although students slightly agreed that the feedback received informed their future writing ability, many had indicated there was a lack of feedback provided, hence partial support was determined. Finally, Warrant 5, “Students act on the test recommendation”, was partially supported, similar to Warrant 1, with approximately 88% of students identified as requiring additional academic essay writing support attending ASC.

5.2 PREDICTING ACADEMIC PERFORMANCE

This section presents results of the academic performance prediction analysis. It responds to the second research question, whether BELA and/or other factors are able to predict academic performance of university students. Specifically, this research question investigated whether BELA scores and/or additional factors (i.e., demographic factors and academic performance data) were able to predict the academic performance of undergraduate students. Academic performance was based on subject assessment and grade outcomes. Concerning academic performance, students were categorised into two groups: 1) students failing (i.e., scoring below 1.00 GPA after two semesters) and 2) students achieving high academic achievement (scoring 3.00 GPA or above after two semesters). Thus, analysis was conducted to predict failure and high achievement at university level. Section 5.2.1 presents

the results of academic performance predictions for students who fail. Section 5.2.2 presents results for students who maintain high academic achievement.

5.2.1 PREDICTING ACADEMIC FAILURE

In order to determine whether BELA or other student demographic and/or academic performance data could predict students failing, binary logistic regression analysis was conducted. Specifically, to estimate the probability of a student failing after two semesters as indicated via GPA of below 1.0 (e.g., .99), which indicated a failing average, a logistic regression was conducted. The student demographics (i.e., gender; age; language background; faculty; indigeneity; disability), as well as academic performance data was added in the model (i.e., overall grade in Core 1; uptake of ASC support; BELA results).

Assumption testing conducted prior to the analysis did not indicate any violation. Overall Core 1: Critical Thinking and Communication grade as well as Major Essay scores were not included in the analysis as they were multicollinear ($r = .87$) and if entered independently, there was a problem of Logit non-linearity. The only other continuous variable (Age) met the assumption of Logit Linearity ($p = .841$). There were a few outliers, but these were kept in the analysis as they were legit cases.

A test of the full model against a constant-only model was statistically significant, χ^2 ($df = 9, N = 1,115$) = 119.08, $p < .001$, indicating that the predictors, as a set, reliably distinguished between whether participants would fail their university degree after two semesters. The entry of the predictors increased the variance accounted for, Cox and Snell $R^2 = .10$, Nagelkerke $R^2 = .18$. The model was 85.7% accurate in its prediction of failing after two semesters. Hosmer and Lemeshow test results confirmed that the model was a good fit for the data, $\chi^2 = 8.36, p = .399$.

Table 5.12 displays regression coefficients, odds ratios, and 95% confidence intervals for odds ratios for all predictors. According to the Wald criterion, gender (male), being a Bond University College (BUC) or a Bond Business School (BBS) student, using EAL/D, being Indigenous, being categorised as having Below satisfactory essay writing via BELA and non-attendance at ASC statistically significantly predicted whether students would fail after two semesters. The odds of failing after two semesters were 4.22 times greater for Indigenous students, 3.88 times greater for students whose writing was assessed as being Below satisfactory via BELA, 3.12 times greater for BUC students, 2.89 times greater for

students who did not attend ASC, 2.02 times for EAL/D students, 1.77 times for males and 1.65 times for BBS students.

Table 5.12

Logistic regression analysis of failing after two semesters as a function of demographics and other academic information (N = 1,115).

	<i>B</i>	<i>SE</i>	Odds Ratio	95% CI for Odds Ratio	
				Lower	Upper
Constant	-4.015***	0.50	0.18		
Gender (Male)	0.57**	0.20	1.77	1.20	2.58
Age	0.10	0.19	1.01	0.97	1.05
Bond University College (Yes)	1.17***	0.23	3.12	2.04	5.09
BBS (Yes)	0.50*	0.23	1.65	1.05	2.59
English Speaker (No)	0.70***	0.21	2.02	1.35	3.02
Indigenous (Yes)	1.44**	0.45	4.22	1.73	10.27
Disability (Yes)	-1.30	1.04	0.27	0.035	2.10
BELA (Below satisfactory)	1.36***	0.27	3.88	2.28	6.61
ASC (Did not attend)	1.05***	0.26	2.86	1.70	4.80

Note. CI = confidence interval; (X) dichotomous variable. * $p < .05$, ** $p < .01$, *** $p < .001$.

5.2.2 PREDICTING HIGH ACADEMIC ACHIEVEMENT

To determine if BELA, other student demographic and/or academic data could predict high academic achievement, defined as a Distinction (i.e., 75%) average or above, a second binary logistic regression analysis was conducted. GPA 3.0 and above was used to indicate high achievement. Variables and assumptions checking were again conducted.

A test of the full model against a constant-only model was statistically significant, χ^2 ($df = 10, N = 1,115$) = 144.33, $p < .001$, indicating that the predictors, as a set, reliably distinguished between whether participants would obtain a GPA of 3.0 or more after two semesters. The entry of the predictors increased the variance accounted for, Cox and Snell $R^2 = .12$, Nagelkerke $R^2 = .19$. The model was 79.2% accurate in its prediction of obtain a Distinction or higher after two semesters. Hosmer and Lemeshow test results confirmed that the model was a good fit for the data, $\chi^2 = 13.10, p = .109$.

Table 5.13 shows regression coefficients, odds ratios, and 95% confidence intervals for odds ratios for all predictors. According to the Wald criterion, gender (female), age (older), English-speaking background, Health Sciences and Medicine (HSM) students, and obtaining a Satisfactory score on BELA statistically significantly predicted whether students would obtain a Distinction or more after two semesters. The odds of obtaining a GPA above 3.0 were 5.16 times for students obtaining a Satisfactory score on BELA, 4.63 times greater for students enrolled in HSM, 1.67 times greater for English-speaking background students, 1.45 times greater for females, and 1.03 times greater for older students.

Table 5.13

Logistic regression analysis of obtaining a Distinction average after two semesters as a function of demographics and other academic information (N = 1,115)

	B	SE	Odds Ratio	95% CI for Odds Ratio	
				Lower	Upper
Constant	-4.31**	0.52	0.01		
Gender (Female)	0.37*	0.16	1.45	1.06	1.98
Age	0.03*	0.02	1.03	1.06	1.98
FSD (Yes)	0.10	0.19	1.11	0.77	1.60
LAW (Yes)	-1.34	0.31	0.72	0.39	1.31
HSM (Yes)	1.53***	0.21	4.63	3.05	7.02
English Speaker (Yes)	0.51*	0.22	1.67	1.09	2.54
Indigenous (Yes)	20.29	7112.99	0	0	0
Disability (Yes)	0.10	0.52	1.11	0.40	3.08

BELA (Satisfactory)	1.64***	0.34	5.16	2.64	10.10
SLS (Did attend)	0.35	0.19	1.42	0.98	2.05

Note. CI = confidence interval; (X) dichotomous variable. ** $p < .01$, *** $p < .001$.

5.2.3 SUMMARY OF RQ2

The main aim of the current study was to determine what factors, if any, could predict academic performance of undergraduate university students at an Australian university. Answering RQ2 achieved this aim. In terms of predicting academic failure after two semesters of university (i.e., GPA below 1.0), it was found that the following characteristics were predictors: gender (male), enrolment in a BUC or a BBS degree, coming from an EAL/D background, being an Indigenous Australian student, scoring 3 or 5 on BELA (i.e., having Below satisfactory essay writing skills) and not attending ASC. The odds of failing after two semesters were 4.22 times greater for Indigenous students, 3.88 times greater for students whose writing was assessed as being Below satisfactory via BELA, 3.12 times greater for BUC students, 2.89 times greater for students who did not attend ASC, 2.02 times for EAL/D students, 1.77 times for males and 1.65 times for BBS students.

In terms of predicting high achievement at university for undergraduate students, defined as achieving a Distinction average (i.e., GPA of 3.0 or above) after two semesters, the following characteristics were found to be predictive: gender (female), age (older), enrolment in an HSM degree, coming from an English-speaking background, obtaining a Satisfactory score on BELA (scores of 7 or 9). The odds of obtaining a GPA above 3.0 after two semesters were 5.16 times for students obtaining a Satisfactory score on BELA, 4.63 times greater for students enrolled in HSM, 1.67 times greater for English-speaking background students, 1.45 times greater for females, and 1.03 times greater for older students.

5.3 CONCLUSION

The main aim of this thesis was to determine what factors could predict the academic performance of undergraduate students at an Australian university, Bond University. A secondary aim was to present a validity argument for BELA, as results on BELA were included in the predictive modelling. This chapter synthesised results of analyses applied in order to respond to the study's research questions.

The first research question was whether a satisfactory validity argument could be presented for BELA. Answering this question required responding to five sub-questions. For RQ1a, support was found for warrants 1, 4 and 5. Partial support was found for Warrant 3. More concerning was a lack of support found for Warrant 2, raters can implement scoring procedures consistently. This requires further discussion. Regarding RQ1b, support was demonstrated for warrants 1, 2 and 3, with partial support found for Warrant 4, test administration conditions are consistent. The concern regarding the fourth warrant is in terms of academic integrity, necessitating further investigation. For RQ1c, support was presented for warrants 1, 2, 3, 4 and 6. No data was collected for Warrant 5. Concerning RQ1d, support was shown for warrants 1, 4 and 5, with partial support found for warrants 3 and 6. Support was not determined for Warrant 2, the test results include feedback on test performance and a recommendation. This is concerning, hence requires further consideration. Finally, for RQ1e, support was found for warrants 1, 2, 6, 7 and 8. Partial support was found for Warrants 4 and 5.

In terms of the study's main aim, the second research question investigated was whether BELA and/or other factors could predict academic performance of university students. Overall, the answer was affirmative. Academic failure was able to be predicted after two semesters of university (i.e., GPA below 1.0), with the following characteristics found to be predictors: male, enrolment in a Bond University College (BUC) or a Bond Business School (BBS) degree, speaking English as an additional language or dialect, being Indigenous Australian, scoring 3 or 5 on BELA (i.e., having Below satisfactory essay writing skills) and not attending ASC. Notably, the odds of failing after two semesters were 4.22 times greater for Indigenous students, 3.88 times greater for students whose writing was assessed as being Below satisfactory via BELA, 3.12 times greater for BUC students, 2.89 times greater for students who did not attend ASC, 2.02 times for EAL/D students, 1.77 times for males and 1.65 times for BBS students.

Additionally, high achievement, defined as maintaining a Distinction average (i.e., GPA of 3.0 or above) after two semesters was able to be predicted. Here, the following characteristics were found to be predictors: gender (female), speaking English as a first language, enrolment in a Health Sciences and Medicine (HSM) degree, and having Satisfactory academic essay writing skills (scores of 7 or 9) as measured by BELA. The odds of obtaining a GPA above 3.0 after two semesters were 5.16 times for students obtaining a

Satisfactory score on BELA, 4.63 times greater for students enrolled in HSM, 1.67 times greater for English-speaking background students, 1.45 times greater for females, and 1.03 times greater for older students. Answering RQ2 achieved the main aim of the current study. The following chapter will discuss these results including their implications.

CHAPTER 6: DISCUSSION

6.0 INTRODUCTION

The main aim of this thesis was to determine what, if any, factors could be used as predictors of academic performance amongst undergraduate students. As BELA results were used in the prediction model, it was necessary to firstly present a validity argument for the assessment. This chapter discusses the results of the analyses in response to the study's research questions. The first research question was whether a satisfactory validity argument could be presented for BELA. The second research question was whether BELA and/or other factors were able to predict academic performance of university students. The chapter firstly discusses findings regarding the validity argument for BELA and then predicting academic performance. The chapter then presents the implications of the study's findings, as well as limitations and future research opportunities. Finally, this chapter provides a conclusion to the thesis.

6.1 SUMMARY OF FINDINGS

This section summarises the main findings in relation to the research questions. It is divided into two sections; section 6.1.1 summarises the validity argument findings for BELA, while section 6.1.2 summarises the academic performance prediction findings.

6.1.1 VALIDITY ARGUMENT FOR BELA

The first research question posed was whether a satisfactory validity argument could be presented for BELA. Presenting a validity argument for the BELA involved determining whether: 1) BELA scores are adequate reflections of observed behaviours, 2) BELA yields results consistent across assessment contexts, 3) BELA provides information on students' skills, knowledge and characteristics consistent with what is understood regarding academic language proficiency, and BELA is an adequate proxy for tasks performed in the academic domain, 4) decisions based on BELA scores are appropriate and well communicated, and 5) consequences of using BELA and decisions informed by BELA are beneficial for all stakeholders. Overall, a satisfactory validity argument was presented for BELA and the use of its scores in the Bond University context. However, concerns threatening the validity of BELA and its uses were identified requiring further discussion. The next sections discuss the

findings of each inference; identified concerns are then considered in relation to extant literature.

The Evaluation inference was mostly supported; however, one serious concern was identified. Regarding supportive evidence, BELA's scoring criteria and rubric were found to capture relevant aspects of academic essay writing (Warrant 1). Academic staff, BELA users and external raters expressed that the three criteria (i.e., Organisation, Linking and flow, and Grammar and vocabulary) were relevant for the task, consistent with findings regarding other PELAs, such as the University of Sydney's MASUS. For MASUS, because the rating criteria and rubrics were established based on language expertise (i.e., Halliday, 1985), relevant components of academic writing are measured (Knoch & Elder, 2016). Concerning BELA, an external rater suggested the inclusion of an additional criterion to assess whether students adequately respond to the task (e.g., "Task response"). However, another argued this was unnecessary. Additionally, the 2-point scale used, comprising Below satisfactory and Satisfactory, eliminated the central tendency effect, described by Landy and Farr (1983, p. 63) as "the avoidance of extreme ... ratings or a preponderance of ratings at or near the scale midpoint". This is advantageous in the context of screening tests, as it forces raters to make unambiguous decisions. The instructions and tasks for BELA were clear for students completing the task (Warrant 4), with students and academic staff in agreement. This finding is similar to that concerning the University of Melbourne's PAAL as evidenced through questionnaire results from two trials (Knoch et al., 2016). Furthermore, BELA was found to be appropriately pitched in terms of difficulty and able to discriminate between more and less able students (Warrant 5). Statistical analysis of test difficulty and test discrimination and feedback from external raters supported this statement. Here, there is similarity with the University of Auckland's DELNA (Elder & Erlam, 2001; Elder & von Randow, 2008), found to be "easy for most candidates" (Read, 2015a, p. 206), yet demonstrating good item discrimination and acceptable reliability.

One warrant in the Evaluation inference was only partially supported, the statement that test administration conditions of BELA were clear and appropriate (Warrant 3). Students indicated they completed the online task in a quiet environment without technical difficulties, and the 60-minute time limit was sufficient. Doubts were, however, raised regarding academic integrity. Students indicated they had not sought help from external sources, yet academic staff expressed concern. On this point, there is similarity with MASUS. Test

administration conditions vary considerably in order to achieve flexibility; however, in certain contexts, such variation may be inappropriate (Knoch & Elder, 2016). With the evolution of Generative Artificial Intelligence, academic integrity concerns are predicted to increase for PELAs, particularly given the software is already being used to co-produce academic work (e.g., Cotton et al., 2023; Kung et al., 2022). Such advancements in technology appear to present as both a threat and opportunity in the domain of language assessment. In terms of obtaining samples of students' "real" work, much of the literature highlights the problem of convincing both EAL/D and English-speaking background students to complete assessments to the best of their ability, without external influence. Perhaps if the beneficial purposes of BELA are promoted in a meaningful way to students, as posited by Read (2008), they may be more fully aware of the task's intentions and academic integrity concerns may be reduced. Setting BELA as a compulsory task to be completed in students' own time may need to be reconsidered, in preference for invigilated completions in a computer laboratory. This was not possible due to budgetary and resource restraints, but it may be a necessary solution given developments in Generative AI.

A more serious threat to the validity of BELA was the lack of support for Warrant 2, that raters could implement consistent scoring procedures. Simply, inter-rater reliability, one of the most important factors in language assessment (Stemler, 2004), was lower than desirable, and while a level of disagreement is unsurprising and tolerable, there are limits (O'Hagan & Wigglesworth, 2015). Inter-rater reliability for some, not all academic staff, was below what is typically reported in the extant PELA literature. For instance, Read (2015, p. 207) described correlations for raters of DELNA ranging from an "unsatisfactory .79 to a highly acceptable .96". In DELNA's favour, although reliability analyses were not routinely conducted, the writing component is "double rated" by two raters, with face-to-face rater training sessions held at the beginning of the semester for all and additional training sessions for new raters (Read, 2015a). An online rater training program was also developed to complement the face-to-face rater training sessions, and three empirical studies have presented the efficacy of the training (Elder et al., 2005; Elder, Barkhuizen et al., 2007; Knoch et al., 2007). At the University of Melbourne, Knoch et al. (2016) reported high correlations (.88) for the Writing subtest during the test's trial. In contrast, each BELA essay is rated by one academic staff member. Having domain experts assess language skills, as opposed to language experts, may explain the variance in ratings. Similar findings have been reported (Elder, Bright et al., 2007; O'Hagan & Wigglesworth, 2015; Smith, 2003), with

ratings by domain experts and language teaching experts differing “widely from one another in the decisions made” (Elder, Bright et al., 2007, p. 38). Academic staff played a critical role in the development and ongoing rating of BELA, and domain experts’ input is highly regarded (Knoch et al., 2020). Without such support of academics, it would be incredibly difficult to present a validity argument for BELA. Nevertheless, domain experts tend to be more lenient when it comes to ratings of essays compared to language assessors (Elder, Bright et al., 2007) due to political, ethical and practical considerations. On this point, the context of BELA is similar to MASUS. Domain experts who rate MASUS essays “are not necessarily well-equipped to make judgements” about linguistic features of texts (Knoch & Elder, 2016, p. 221). Thus, for BELA and MASUS, it is concerning that determinations of who is at risk could be somewhat random, and fairness could, consequently, be questionable. However, differences in the scoring between raters, whether domain experts or language experts, can be derived from numerous sources (Morgan et al., 2014), and it is highly unlikely that perfect agreement amongst writing raters can be achieved, regardless of expertise (Knoch & Elder, 2016; Knoch, Fairbairn et al., 2018).

A possible solution to enhance levels of consistency of raters is to provide more in-depth and frequent rater training whilst also monitoring ratings. When sufficient and effective rater training occurs, it is possible for single raters to rate student essays with increased accuracy (Morgan et al., 2014). Training allows raters to be re-introduced to the criteria and may assist in focusing their attention on all aspects of the rating scale, as opposed to focusing on a particular component of writing (Shohamy et al., 1992). Evidence has found that training can eliminate extreme differences in rater severity, enhance the self-consistency of raters and limit bias (McIntyre, 1993; Lumley, 2002; Weigle, 1998). However, it is difficult to allocate time for raters to collaborate and synchronously moderate their ratings in traditional, face to face rater training (Brown & Jaquith 2011; Knoch, Fairbairn et al., 2018). Moreover, face to face training may be intimidating with raters requiring differing time allocations to adequately rate written scripts (Charney, 1984; Elder, Barkhuizen et al., 2007; Hamp-Lyons, 2007). Online training, therefore, may be an effective mode of rater training.

Online rater training allows raters to access training materials and sample writing scripts anywhere, anytime, which in turn enables them to re-orientate themselves to the process of rating, including the rating scale and monitor their own rating behaviour (Brown & Jaquith, 2011; Elder, Barkhuizen et al., 2007, Elder et al., 2005). Studies have demonstrated

that online rater training is just as effective as in-person training (Elder, Barkhuizen et al., 2007; Knoch et al., 2007; Knoch et al., 2018) and favourable amongst raters (Elder, Barkhuizen et al., 2007; Knoch et al., 2007). Consequently, it is recommended that an online rater training module be created for continuing and new BELA raters. It is also important to monitor raters' on-going scoring to identify discrepancies and offer support. Monitoring ratings may allow for the identification of inconsistencies (Harsch & Martin, 2012) and assist raters in understanding their decision-making (Mumford & Atay, 2021).

6.1.1.2 GENERALISABILITY INFERENCE

The Generalisability inference was supported. The two versions of BELA were found to be parallel in design (Warrant 1). Although versions are not trialled as extensively in comparison to other PELAs (e.g., Read, 2015a), BELA was piloted prior to official implementation (Lydster & Brown, 2017). Objective review by external raters also noted their parallelism. The BELA process was found to use appropriate equating procedures to ensure the two versions comprise equivalent difficulty (Warrant 2). Analyses indicated no statistically significant difference in scores between version 1 and version 2. If additional versions are added to the pool of BELA tasks, as recommended by one external rater in response to academic integrity concerns, it is necessary to evaluate them to ensure equivalency and that they elicit adequate samples of test taker writing ability (Knoch et al., 2016). Support was also found for BELA having sufficient tasks that provide stable estimates of students' ability (Warrant 3). Pilot study results were compared with the current study's data, and findings revealed BELA provides stable estimates of test taker ability. Although BELA is a single writing task with a time limit of 60 minutes, it is argued that it would be difficult to convince students to complete any additional writing tasks for such low-stakes purposes; thus, extending the time limitation would be ill-advised when 60 minutes is considered the upper limit (Knoch & Elder, 2016).

Only partial support was found for Warrant 4 of the Generalisability inference, "test administration conditions are consistent." The nature of BELA being an online, self-administered assessment means it is difficult to determine whether administration conditions are indeed consistent. PAAL is similar in this regard, "as students can take the test in their own time at a place of their choosing, it is likely that the conditions are not absolutely consistent" (Knoch & Elder, 2016, p. 33). However, as PAAL is a low stakes test, the conditions it is completed in "are probably not of great concern" (Knoch et al., 2016, p. 33).

For BELA, as noted in response to Warrant 3 of the Evaluation inference, students indicated they completed the assessment in a quiet environment, the time limit was appropriate, and they had no technical issues. Students also indicated they had not plagiarised; however, staff stakeholders voiced concerns regarding academic integrity. As noted in the previous section, further investigation is required concerning academic integrity and BELA; however, the Generalisability inference was supported.

6.1.1.3 EXPLANATION AND EXTRAPOLATION INFERENCE

The Explanation and Extrapolation inference was supported. Findings revealed that BELA results are good predictors of language performance in the academic domain (Warrant 1), with Point Biserial correlations indicating that obtaining a satisfactory score on BELA (i.e., 7 or 9) explained 9.6%, 14.4%, and 11.6% of the variance for the Major Essay, overall grade and GPA respectively, after two semesters. In other PELA contexts, studies have demonstrated higher levels of variance between PELA results and academic performance. For instance, DELNA was shown to demonstrate a moderate to high relationship of 30.3% for overall grade and 10.2% for GPA when examining academic performance amongst Bachelor of Business and Bachelor of International Management students (Elder & Erlam, 2001). However, in a more comprehensive study by Elder, Bright et al. (2007), weaker, yet significant relationships of 5.3% and 6.3% between DELNA scores and students' first semester GPA were found. Regarding MASUS, in a predictive study using scores on the PELA and assignment results in the UK, the PELA was found to explain 60.1% of the variance on a Social Sciences subject assessment (Erling & Richardson, 2010). However, in the same study, the variance was found to be 10.2% in an Information and Communication Technology subject. Typically, correlations between PELA scores and academic performance of approximately 0.30 to 0.40 have been determined. In terms of variance, this equates to between 9% and 16% of students' academic performance being explained by their PELA results (Elder, Bright et al., 2007). Hence, the results concerning BELA are in accordance with the existing literature. Support was also found for BELA task characteristics being similar to undergraduate assessments (Warrant 2). This was supported by student and academic stakeholders and analysis of common assessment genres at the research site. The developers of BELA, including the researcher, drew on their English for academic purposes and language assessment experience, which further strengthens support for this warrant, consistent with Knoch et al.'s (2016) review of PAAL.

Furthermore, the current study found that BELA scores provided sufficient information about students' academic language proficiency (Warrant 4). Students and academic staff agreed that BELA was indicative of students' academic writing skills and external raters expressed that BELA provided adequate insight about students' capabilities. DELA and MASUS "model their tasks on the academic domain" (Knoch & Elder, 2016, p. 223), and BELA is similar in this regard. Like MASUS, BELA may be considered to underrepresent the overall academic domain, as the task assesses only one skill, academic essay writing. However, referring to the purpose of BELA, it is argued that BELA provides adequate information about students' academic essay writing, a specific and important element of academic language proficiency. In addition, BELA tasks were demonstrated to not unfairly favour certain groups of students (Warrant 6). Students, in particular, expressed the assessment was fair. Other stakeholders including staff and external raters indicated the task was fair and did not disadvantage particular groups of students. In other contexts, this warrant was seen by Read (2015, p. 213) as "potentially of concern to DELNA" due to the inclusion of both English-speaking background and EAL/D students completing the assessment, with Elder and Erlam (2001) determining bias to be an issue when DELNA was first introduced. In BELA's case, there was a higher representation of EAL/D students in the Below satisfactory category, yet it is argued that this finding would be expected. Further investigation, in accordance with Elder et al.'s (2003) suggested methodology, may determine whether bias, and therefore fairness, is a factor. This may include employing sophisticated methods of analysis, such as Rasch modelling, to determine precisely if and where bias occurs (Fan et al., 2019; Fan & Knoch, 2019).

Partial support was found for Warrant 3, "Linguistic knowledge, processes, and strategies employed by test takers are in line with theoretically informed expectations and observations of what is required in the corresponding academic context". There was no definitive answer in terms of whether students employed typically expected practices in completing BELA. Many students stated they brainstormed ideas and planned prior to commencing, yet results were inconclusive amongst academic staff. Importantly, both groups concurred that language used in BELA was consistent with language required at tertiary level. In other university PELA contexts, such as DELNA (Read, 2015a) and PAAL (Knoch & Elder, 2016), this type of insight has not been obtained.

Within the Explanation and Extrapolation inference, the only warrant that was not investigated was Warrant 5, students' performance on BELA would be related to their performance on other assessments of academic language proficiency, such as IELTS. Given that PELAs and high-stakes language exams are disparate, particularly regarding their purposes, it is unsurprising that, to date, no published study has investigated the relationship between a PELA and a pre-entry exam. In the DELNA context, Read (2015) explained the decision was made not to use high stakes language test scores as a measure to compare with DELNA, as the PELA is administered to English-speaking background students, who may not be required to complete a language assessment for university entry, as well as EAL/D students. Instead, the University of Auckland chose to use grades in academic subjects as the criterion measure to compare DELNA scores (Smith, as cited in Elder, Bright et al., 2007). Read also suggested an alternate method may be to obtain academics' views of the written piece and how they would assess it. The latter point may be worthy of investigation in future studies.

6.1.1.4 DECISIONS INFERENCE

Overall, the Decisions inference was mostly supported; however, a notable concern was identified regarding the provision of feedback. First of all, students were found to be correctly categorised based on their BELA scores (Warrant 1). Analysis of students' results on the Major Essay and subsequent GPA after two semesters, indicated that those scoring low on BELA would, on average, demonstrate lower Major Essay results and GPA; the reverse was observable for students scoring high on BELA. It should be noted that standard setting procedures undertaken at other universities have been more robust. Concerning DELNA, Elder and von Randow (2008) determined a cut score that was sufficiently sensitive to identify 93% of students at risk of academic failure due to low English language proficiency. However, the cut score eventually applied was set "relatively low" due to budgetary concerns (Read, 2015a, p. 212). In the context of PAAL, Knoch et al. (2016) stated that standard setting entailed two processes. The first was during the development of the test, when a Receiver Operating Characteristics curve analysis, a technique for standard setting, was utilised to determine cut-scores on the Screening test. The second process involved a team of trained language assessors setting the cut-scores of the writing subtest through individually rating and then discussing their ratings of 50 writing texts. For BELA, the current study's findings indicate standard setting procedures were adequate. However, due to inconsistency amongst raters, a future study could investigate the outcomes of including those scoring 7 in

the Below satisfactory category. This research opportunity will be discussed further in reference to Warrant 2.

In addition, the results of BELA were found to be distributed in a timely manner to stakeholders (Warrant 4). Students agreed that results were distributed timely, with a significantly stronger positive opinion amongst BELA users. For DELNA, the results of the initial Screening component of the test are automated and sent to students within 24 hours. For the Diagnostic part of the test, due to the writing component being cross-marked, there is a 10-day turnaround for students to receive feedback (Read, 2015a). Regarding PAAL, Knoch et al. (2016) reported that results are distributed within one to two days after the assessment has been completed. Although not as speedy a turnaround as PAAL, BELA's turnaround time was considered adequate by stakeholders. Similarly, results were demonstrated to be available to all relevant stakeholders (Warrant 5). BELA users maintained that all stakeholders were given access to BELA data. In other contexts, results are provided to stakeholders on request (Knoch et al., 2016; Read, 2015a). It is believed that sending BELA data to staff who may influence support uptake or decision making about support increases awareness of the PELA processes and its role within the institute's retention and success policy.

Warrant 3, "the recommendation is closely linked to on-campus support", was only partially supported. Students maintained that academic writing support was available on campus yet had indicated they had not received a recommendation. In contrast, BELA users strongly believed that students were directed to on-campus support. This point concerning a recommendation is connected to Warrant 2 regarding the provision of feedback and is discussed in the next paragraph. Partial support was also determined for Warrant 6, "Test users understand the meaning and intended use of the scores". Staff indicated they had understood BELA scores' meaning. However, students expressed a lack of clarity, in particular, regarding Below satisfactory scores. One solution may be to create a guide to distribute to stakeholders across the university to provide information on BELA, its purposes, and interpretations of scores and their consequences. A BELA reference group may also be established to discuss issues and opportunities, as is done with DELNA at the University of Auckland (Read, 2015a).

A lack of support was evident for Warrant 2, "The test results include feedback on test performance and a recommendation." Specifically, clarity from stakeholders regarding

feedback was lacking. Some students expressed that they received valuable feedback from completing BELA; for instance, one student stated, “I found the experience helpful and the feedback valuable.” However, several students indicated feedback was not provided (e.g., “we received no feedback”, “Could have been beneficial to receive feedback.”). Feedback is considered one of, if not the most, impactful influences on learning (Hattie & Timperley, 2007) and can lead to development in accuracy, especially for writing (Humphreys, 2015; Knoch, 2012; Li & Barnard, 2011; Rummel & Bitchener, 2015). Given that students’ first year experiences are often connected to their first assessments, inconsistent feedback is an important finding. Students performing well on their first assessment feel a sense of belonging, while the opposite may be true for those performing poorly (Walker-Gibbs et al., 2019). A lack of feedback, especially for those scoring Below satisfactory, may result in the unintended consequence of students questioning their place within academia. In other contexts, a lack of feedback has also been identified. Regarding PAAL, student participants stated they were disappointed with their statement of results and called for more detailed, diagnostic analysis of their performance with an advisor (Knoch et al., 2016). Participants had issues with the “vagueness of the support recommendation given in the report, with many of them wanting a clearer directive for what was required of them” (Knoch et al., 2016, p. 37). A similar situation in the DELA context is also apparent (Knoch & Elder, 2016).

A possible explanation for the finding that some students were provided insufficient feedback is that only students categorised as having Below satisfactory are required to attend ASC consultations. During consultations, Learning Advisors provide detailed feedback on students’ essays, and as the text is relatively short (i.e., 300 to 400 words), sufficient feedback can be provided in a 30-minute consultation, with follow-up consultations, workshops and other resources recommended as appropriate. Students in the Satisfactory category are welcome, yet not obliged to attend ASC consultations. It is believed students who indicated there was a lack of feedback were in the Satisfactory category. This context is similar to that of the DELNA. The most detailed feedback is provided to students identified as being most at-risk of academic failure during a 30-minute consultation with a Language Advisor at the DELNA Office, in which the results are reviewed (Read, 2015a; Rummel, 2020). However, students who take the writing test of DELNA and receive a score above the threshold score receive general advice on enhancing their academic ELP (Read, 2015a). This is not provided to students with Satisfactory essay writing skills according to the BELA process and may be a potential solution. A further solution, which would require further investigation, is to ensure

more students receive feedback by strongly recommending those scoring 7 on BELA to attend ASC consultations. As the Satisfactory category includes students scoring 7, this indicates that one area of academic essay writing has been identified as Below satisfactory. Rather than only requiring students scoring 3 and 5 on BELA to attend ASC, it may be beneficial to also encourage students who scored 7 to visit ASC. This may also involve examining a change of cut score threshold and including those scoring 7 in the Below satisfactory category, as suggested in the previous section.

6.1.1.5 CONSEQUENCES INFERENCE

The Consequences inference was supported. Firstly, it was determined that BELA does not result in any stigma or disadvantage for students (Warrant 2), with each stakeholder group believing the chances stigma was attached to poor performance on BELA was low. Students, in particular, strongly disagreed that some may have felt disadvantaged when completing BELA. This is consistent with DELNA policy, for which Read (2015, p. 218) argued, “the fact that university policy does not exempt any students from the assessment reduces, if not eliminates, any sense of stigma associated with taking the DELNA screening.” At the time of writing, discussions were ongoing to integrate BELA into the Bachelor of Medical Studies, which would mean all undergraduate students at Bond University would complete BELA, further strengthening support for this warrant. The current research also found that students’ perceptions of BELA and its usefulness were positive (Warrant 3). All stakeholders noted the utility of the task; some students indicated it was unnecessary, but many of these students believed it benefitted others. An indicative quote was, “The idea is good. I believe it would be beneficial for people who severely lack in an area of academic writing.” There appears to be a paucity of published feedback from test takers, so this finding makes an important contribution. Read (2015) argued that procedurally monitoring students’ opinions and experiences via an anonymous online survey and interviews satisfies this warrant regarding DELNA. In contrast, there were mixed feelings amongst PAAL test takers; students were positive towards the relative ease of completing the task online yet were less appreciative of the feedback afforded to them and the support options available (Knoch et al., 2016).

The follow-up language development options provided to students were considered appropriate (Warrant 6). This was supported by internal stakeholders, who maintained that attending ASC helped improve students’ academic writing. To ensure this continues to be the

case, an English language support taskforce, who present recommendations concerning language policy, could be established as per the University of Auckland context (Read, 2015a). Furthermore, students who engaged in support options were documented to improve their English over the course of their studies (Warrant 7). The majority of students who had completed the original BELA, BELA 2 and attended at least one consultation at ASC demonstrated gains in their results on the second BELA. More students attending ASC showed gains in their writing compared to those who did not attend. This finding is notable, as the development of students' ELP as a consequence of participating in on campus language mechanisms "is seldom explored in PELA contexts" (Knoch & Elder, 2016, p. 226). As most PELAs are one off assessments, it is difficult to ascertain language development (Knoch & Elder, 2013). Studies investigating language gains using high stakes, commercialised assessments, such as IELTS, have found students generally perform better on the assessment later in their studies (Humphreys, 2015; O'Loughlin & Arkoudis, 2009). However, this is not always the case. Humphreys (2015), for example, found a minority of graduates obtained a lower overall IELTS exit score than their entry score. Other studies have been inconclusive (Knoch et al., 2015; O'Loughlin & Arkoudis, 2009; Storch & Hill, 2008; Storch & Tapper, 2009). To the researcher's knowledge, no published data has been collected regarding language development in other PELA contexts with Knoch and Elder (2016, p. 226) acknowledging that "the warrant relating to students improving their language skills as a result of the language support offered on campus cannot be supported."

Finally, students who did not act on BELA recommendations to attend ASC consultations were proven to be more likely to struggle in their studies (Warrant 8). Students who did not attend ASC significantly scored lower in their Major Essay, overall grade in the subject and had significantly lower GPAs after two semesters compared to students who had acted on the recommendation to attend ASC. Non-participation is likely an important issue. It is apparent that compliance rates of PELAs and acting on test recommendations vary considerably at different universities, particularly when the participation is not compulsory (Ransom, 2009). It is argued that those who do not participate in language development mechanisms are often the ones who require it most (Murray, 2012; Podorova, 2016). For instance, Read and von Randow (2013) determined that the majority of students who did not participate in the follow up stage of DELNA had low GPAs. A further gap in the literature exists regarding the academic performance of students who fail to complete PELAs. A previous BELA study found that of the students who elected to not complete the task, 34 out

of 48 students (68.75%) failed the subject, Core 1: Critical thinking and Communication. It would appear that non-completion of a PELA, especially when compulsory, is another early warning sign and should form part of universities' retention and success mechanisms.

In contrast to the warrants that were fully supported, only partial support was found for Warrant 1, "All target test takers sit the test". Approximately 95% of target test takers completed BELA. Although not universal uptake, this finding appears to be positive in comparison to the limited published data available for other PELA contexts. For DELNA, it was estimated that "around 90 per cent" sit the Screening component while "around 60 per cent" comply with the recommendation to complete the Diagnostic test (Read, 2015a, p. 215). PAAL completion rates of students enrolled in two degrees, the Bachelor of Commerce and the Master of Engineering were 35% (110 out of 310 students) and 12% (60 out of 491 students) respectively. Knoch et al. (2016) opined that institutional policy made it impossible to make the PELA compulsory, as "it goes beyond content course requirements" (Knoch et al., 2016, p. 38). In other contexts, such as DELA, many students who are identified as required to complete the task elect not to (Ransom, 2009). Therefore, the goal of all targeted test takers completing the test is not easily achieved. A strength of the BELA process is the near universal uptake rate, which is a result of the PELA being embedded in the curriculum of a core subject worth 2% of students' overall grades.

Similarly, Warrant 4, "The feedback from the test is useful and directly informs future learning", was only partially supported. Students had, on the whole, agreed that the feedback received via BELA informed their future writing skills. Nevertheless, several respondents had indicated that there was a lack of feedback provided (see Warrant 2 of the Decision inference); hence, this warrant could only be partially supported. As explained in section 6.1.1.4, the most detailed feedback is provided to students identified as having Below satisfactory academic essay writing (i.e., scores of 3 or 5). This is a similar situation to what occurs at the University of Auckland (Read, 2015a). Regarding PAAL, Knoch et al. (2016) acknowledged that feedback is minimal, with students expressing disappointment with the statement of results and requesting more detailed diagnostic analysis of their performance. Specifically, students took issue with the "vagueness of the support recommendation given in the report, with many of them wanting a clearer directive for what was required of them" (Knoch et al., 2016, p. 37). Again, this is not dissimilar to the BELA context at Bond University. It may be difficult to provide adequate feedback to each and every student

completing the assessment and, perhaps, beyond the scope of the PELA. Thus, it is believed the lack of support does not pose a serious threat to the assessment's validity.

Finally, Warrant 5, "Students act on the test recommendation", was also partially supported. Similar to Warrant 1, approximately 88% of students identified as requiring additional academic essay writing support attended ASC; hence, universal uptake of support was not achieved. Although not desirable, reaching close to 90% of those who may be at risk is commendable. For DELNA, a recent study by Erlam and Botelho de Magalhaes (2021, p. 41) noted uptake was "a long-standing challenge for DELNA ... and continues to be so." In the second year of the study, 72% of students engaged with the DELNA language advisor compared to 63% in the first year. The researchers highlighted that facilitating online advisory sessions was one reason for the increased participation. Limited data has been presented for other PELA contexts. Concerning PAAL, it was stated that approximately 15% of test takers were categorised as being at risk of academic failure, yet it was unclear how many of these students acted on the recommendation of enrolling in language enhancement courses; however, historical enrolments in such courses were low (Knoch et al., 2016). For MASUS, test taker uptake of support is not mandatory as part of the PELA process, and this data is not routinely tracked (Knoch & Elder, 2016). Throughout the PELA literature, it tends to be the high achieving students who will actively seek out and utilize support mechanisms. In Fox, von Randow et al. (2016), for instance, "only a few students followed-up immediately after receiving their initial diagnosis and the majority were A+ students, who tended to access all and any form of support" (p. 274). Students who require support elect not to for various reasons (von Randow, 2013). Students can be nervous, unaware and/or unwilling to seek support from university services, such as ALL centres (McLeod, 2012). Alternatively, Elder and Erlam (2001) suggested that poor uptake could be due to students mistakenly believing receiving support may in some way jeopardise their progress in their degree programmes. Enforcing compulsory participation is contentious. Some argue that imposing mandatory English language enhancement without imposing sanctions for non-compliance is "pointless"; others feel that it is the role of a university to provide advice rather than impose sanctions (Elder & Read, 2015, p. 34). Overall, the finding that almost 90% of students who were recommended to attend ASC consultations did so, is a notable finding.

6.1.1.6 SUMMARY OF VALIDITY ARGUMENT FINDINGS

Overall, a satisfactory validity argument was presented for BELA, including its scores and subsequent decisions and consequences. Each inference was supported, and BELA and its associated processes are well situated in a context that allows for its integration into a compulsory undergraduate subject, providing extended reach to the majority of students at Bond University. Cohorts of undergraduate students may be screened based on their academic essay writing skills and those who may benefit are provided with effective on-campus support. The University is to be commended for the provision of financial and human resources to support its students.

There are concerns with certain aspects of the assessment, some considered serious, which require further investigation and for which recommendations have been suggested. The main discussion points here were concerns regarding academic integrity. It is anticipated that, with the evolution of Generative Artificial Intelligence, concerns will become more prevalent. Moreover, inter-rater reliability analysis demonstrated a lack of adequate consistency between raters. This point is considered the main threat to the validity argument for BELA. Consequently, it was recommended that more in-depth and frequent rater training sessions be delivered to BELA raters, online. Raters' scoring should also be monitored to identify any discrepancies. Furthermore, students maintained that academic writing support was available on campus, yet some indicated they had not received a recommendation. In regard to BELA scores and users' understanding of their meaning and use, students expressed a lack of clarity regarding Below satisfactory scores, in particular. Several students indicated they received valuable feedback, yet multiple students explained that feedback was not provided. A possible explanation for this finding is that only students whose writing was categorised as Below satisfactory were required to attend an ASC consultation to receive feedback. Students in the Satisfactory category were welcome, yet not obliged. This is a possible threat to the validity of the PELA; however, it may be difficult to provide in-depth feedback to each student and, perhaps, beyond the scope of BELA.

Approximately 95% of target test takers completed BELA, and although this was not universal uptake, is a commendable achievement in the context of PELAs. This is similar to the finding regarding uptake of support, with approximately 88% of students identified as requiring additional academic essay writing support attending ASC. Increasing the promotion of the assessment to highlight its benefits is recommended to increase participation in the

assessment and feedback processes and, perhaps, reduce academic integrity concerns. Having presented an adequate validity argument for BELA, the next section examines factors able to predict academic performance of undergraduate students.

6.1.2 PREDICTING ACADEMIC PERFORMANCE

This section discusses the findings concerning academic performance predictions. It focuses on the main of the thesis and responding to the second research question, whether BELA and/or other factors could predict academic performance of university students. Section 6.1.2.1 presents the findings of performance predictions for students who fail their undergraduate studies (i.e., obtaining a GPA below 1.00 after two semesters). Section 6.1.2.2 presents findings for students who maintain high academic achievement, defined as achieving a Distinction average (i.e., 75%) or above during their studies (achieving a GPA of 3.00+ after two semesters).

6.1.2.1 PREDICTING ACADEMIC FAILURE

The first analysis investigated which factors predicted student failure (i.e., obtaining a GPA of .99 and below) after two semesters of study. This study found the following characteristics to be predictors: Indigeneity (Indigenous Australian), Below satisfactory academic writing skills, language background (EAL/D), non-attendance at ASC, enrolment in Bond University College (BUC) or Bond Business School (BBS), and gender (male). The odds of failing after two semesters were: 4.06 times greater for Indigenous Australian students compared to non-Indigenous students, 3.77 times greater for students with Below satisfactory academic writing as measured by BELA compared to students with Satisfactory writing, 3.12 times greater for BUC students compared to students from other faculties, 2.89 times greater for students who did not attend ASC compared to those who did, 1.93 times greater for males compared to females, 1.89 times for EAL/D students compared to English-speaking background students, and 1.74 times greater for BBS students compared to students from other faculties. The next sections discuss each factor.

6.1.2.1.1 INDIGENOUS STUDENTS

Undergraduate Indigenous Australian students were found to be 4.06 times more likely to fail compared to non-Indigenous students. Throughout the literature, concerns have been identified regarding both retention and success rates amongst Indigenous students at university (Anderson et al., 2022; Bradley et al., 2008; Rochecouste et al., 2016; Universities

Australia, 2017). Universities Australia (2020) aimed to increase Indigenous student participation and graduation, setting the goal to achieve equal retention and success rates amongst Indigenous students and their non-Indigenous counterparts by 2025. However, according to data provided by the Tertiary Collection of Student Information (Department of Education, 2023b), the attrition rate amongst Indigenous, bachelor degree students was 22.52% compared to 12.12% for non-Indigenous students in 2020. At the sub-bachelor level, including diploma programmes, these figures were 46.44% and 30.63% respectively. In terms of retention, the rate for Indigenous bachelor degree students was 77.30% in comparison to 87.55% amongst non-Indigenous students. For sub-bachelor programmes, the retention rates were 52.27% and 67.96% for Indigenous and non-Indigenous students respectively. In regard to success, in 2021, the success rate for Indigenous, bachelor degree students was 72.81% compared to 85.78% amongst non-Indigenous students. At the sub-bachelor level, these figures were 58.23% and 77.23% respectively. Consequently, the goals set by Universities Australia are noble and aspirational but fall short of desired targets, and time is of the essence to achieve parity in higher education.

At the research site, in 2020, the attrition rate for Indigenous bachelor degree students was 22.52%, approximately 10% higher compared to non-Indigenous bachelor degree students with 12.53% (Department of Education, 2023b). For Indigenous students enrolled in sub-bachelor programmes, including diplomas, the attrition rate in 2020 was 46.44% compared to 33.99% for non-Indigenous students. The retention rate for bachelor degree Indigenous students was closer to parity at 77.30% compared to 87.29% for non-Indigenous students. For sub-bachelor degree students, these figures were 52.27% for Indigenous students and 64.51% for non-Indigenous students (Department of Education, 2023b). Finally, the success rate for Indigenous students in 2021 was 72.81% for bachelor degree students and 85.17% for non-Indigenous students. For sub-bachelor programme students, success rates were 58.23% for Indigenous students and 75.88% for non-Indigenous students (Department of Education, 2023b). These statistics are notable, as Indigenous students not achieving their academic goals may impact decisions made by others in the community regarding whether to commence a degree programme (Koramannil, 2016). Hence, providing Indigenous students with the opportunities, knowledge, and skills which may result in academic success is vital.

A meta-synthesis of the challenges faced by Aboriginal and Torres Strait Islander students at university was conducted by Anderson et al. (2020). Overall, the common themes

identified that Indigenous students may suffer from social isolation, especially with non-Indigenous counterparts, and face a lack of academic confidence. Despite intrinsic motivation (Anderson et al., 2022), there may be a lack of understanding around the expectations and workload for those who are first in family (Hill et al., 2020), as well as a lack of preparedness for university (Lydster & Murray, 2018). This is the case for many students, regardless of indigeneity. Although other important considerations were identified by Anderson et al.'s (2022) analysis including a shortage of Indigenous academics and Indigenous knowledges embedded within curricula, contributing to academic confidence is an area in which the BELA process may assist. If Indigenous students are identified as requiring assistance with the conventions of academic essay writing, for example, there is an opportunity to collaborate with the Indigenous Centre of the University and Indigenous Support Officers, commonly referred to as Indigenous Tutoring Assistance Scheme (ITAS) tutors, to assist in building students' confidence. ITAS has been identified as playing a key role in Indigenous students achieving academic success (Lydster & Murray, 2019; Whatman et al., 2008). The provision of such tutoring schemes results in increased confidence, lowered stress and higher academic performance (i.e., grades) amongst the Indigenous student participants. Hence, it is argued that further collaboration between support units is worthy of exploration.

It is critical, however, that support for Indigenous students does not comprise a deficit model, whereby assistance is proffered based on the invalid premise that they commence studies ill-equipped compared to their non-Indigenous counterparts (Buxton, 2017; Nakata et al., 2008; Nakata et al., 2018; Syron & McLaughlin, 2010; Whatman et al., 2008). Further, research involving Indigenous peoples and their communities must be beneficial for them (DiGregorio et al., 2000). Therefore, it is imperative that the current finding does not merely add more statistics to the data already published on the gap between Indigenous and non-Indigenous students. Instead, the current study must result in beneficial outcomes for Indigenous students at the University. An opportunity exists to act on this finding so that Indigenous students' academic confidence is enhanced, and success is celebrated.

6.1.2.1.2 BELOW SATISFACTORY WRITING AS MEASURED BY BELA

This study determined that students scoring 3 or 5 on BELA, thus being categorised as having Below satisfactory academic essay writing, were 3.88 times as likely to be failing after two semesters compared to students having Satisfactory writing. This finding is consistent with a previous study investigating the relationship between BELA scores and

academic performance (Lydster & Brown, 2017). Lydster and Brown examined the extent to which students scoring Below satisfactory on BELA were at risk of failing the Major Essay. Overall, the lower a student's score on BELA, the higher the absolute and relative risk the student would fail the Major Essay. For students who scored 3 on BELA, 43.8% failed the Major Essay. In terms of risk relative to students who scored 9 on BELA, these students were 4.78 times more likely to fail the Major Essay.

To date, there has been a paucity of empirical studies investigating the ability of a PELA to predict academic performance at university. The PELA that has received the most attention has been the University of Auckland's DELNA. In a recent study, Erlam and Botelho de Magalhaes (2021) analysed the correlation between students' DELNA results and their GPAs in the same year they completed the Screening component of the DELNA. Examining the data of 5,469 undergraduate students, a statistically significant, modest correlation was determined (.33) between results on DELNA Screening component and GPA. The GPA outcomes of students identified as having below satisfactory academic language skills were skewed towards achieving lower GPAs, suggesting this group of students are at increased risk of failing their subjects in their first semester. The same analysis was conducted for students scoring above the threshold score with this group being more inclined to achieve higher GPA outcomes in comparison to students scoring below the threshold.

Further studies have demonstrated the ability of practical, low stakes and simple to administer PELAs to predict academic performance. Daller and Wang (2017) determined that a 25-minute C-Test, in which half of every second word was omitted, and 30-minute writing task were better predictors of academic performance of a sample of 107 mostly Chinese students at Swansea University in the UK than IELTS. The C-Test was found to predict 21% of students' GPA. The researchers concluded that general language proficiency and vocabulary knowledge were "two key predictors of study success" (Daller & Wang, 2017, p. 368). Furthermore, a Belgian study analysing the performance of students on a post-entry, online reading and vocabulary screening test found that results correlated with "credit completion rate". The correlation of approximately .30 was described as "a small but meaningful predictor of achievement" (Heeren et al., 2021, p. 1). The researchers concluded that the low stakes, practical PELA was useful as an early detector of students who may require further diagnosis or support to develop academic language skills in Dutch. Studies finding a relationship between PELA scores and academic performance are important, as it

has been found that low ELP is a serious concern (e.g., Birrell et al., 2006). At the research site, both international EAL/D and domestic students have expressed that academic writing genres, such as critical research papers, were problematic (Webb, 2011). The current research, therefore, highlights the importance of developing academic writing skills early in students' degrees.

The current research's findings are notable, as they demonstrate the important role academic language plays at university. It is necessary to highlight that students scoring low on BELA (i.e., Below satisfactory) included both EAL/D and English-speaking background students. Consequently, it is not just "an international student problem", supporting findings by Botha (2022) that "academic communication skills are a problem for a surprisingly large proportion of English-speaking Australian students" (p. 1). This research also supports the value of identifying students whose academic writing skills are below standard early in their first semester of study. Once identified, these students are put in touch with Learning Advisors at ASC who can provide detailed feedback on students' writing and offer recommendations for development. A PELA will never be a silver bullet and will not assist students in developing their language skills. It is also important to stress that students' communications skills are not guaranteed to develop, unless students act or engage with support mechanisms.

6.1.2.1.3 EAL/D STUDENTS

A third predictor of academic failure was language background. Specifically, EAL/D students were 2.02 times more likely to fail after two semesters in comparison to students from English-speaking backgrounds. Language background has been found to play a role in academic performance predictions, yet the findings are inconsistent. Multiple studies have found that EAL/D students perform worse academically compared to English-speaking background students (Bougnan, 1993; Gerardi, 1996; Grebennikov & Skaines, 2009; Smith et al., 2012). This has been found regarding nursing (Bougnan, 1993; Gerardi, 1996) and accountancy (Smith et al., 2012), for instance. The latter study, involving a first-year accountancy subject at Edith Cowen University, determined that language background was the most important variable in distinguishing between students who succeeded and students who failed the subject (Smith et al., 2012). Furthermore, in a study of undergraduate students at the University of Western Sydney, EAL/D status was found to be associated with low GPA (Grebennikov & Skaines, 2009). It is interesting to note the focus on poor performance,

indicating that studies have focused on failure and/or attrition, not necessarily retention and success.

In contrast, studies have also found language background plays only a minimal role in terms of academic performance (Berman & Cheng, 2010; Wu Pong et al., 1997). A Canadian study by Berman and Cheng (2010) examined perceived difficulty of various language skills amongst EAL/D and English-speaking background students, as well as academic outcomes. It was found that EAL/D students perceived language skills to be more difficult in their studies. However, in terms of actual performance, undergraduate EAL/D students achieved similar results in their studies compared to their peers. Other studies have demonstrated that language status was unrelated to GPA (Wu Pong et al., 1997). The study by Cao and Gabb (2006), examining 12,500 commencing students and their characteristics at Victoria University found similar attrition rates for both EAL/D and English-speaking background students. Notably, EAL/D students have been found to demonstrate lower rates of attrition than their English-speaking background student counterparts (McMillan, 2005), and this is supported by recent national data. Regarding retention, the data provided by the Tertiary Collection of Student Information (Department of Education, 2023b), indicates retention rate for EAL/D students, termed “non-English speaking background”, at bachelor degree level was 91.79%, slightly higher than their “not from non-English speaking background” peers with 86.88% in 2020. At the sub-bachelor level, including diploma programmes, these figures were 74.41% and 63.80% respectively. These data are consistent looking back at pre-Covid-19 years’ data to 2015 (Department of Education, 2023b), indicating EAL/D students were more inclined to continue in their degree programmes than English-speaking background students. There are inconsistencies amongst the literature, and the finding that EAL/D were more inclined to fail in the current study appears to go against the trend found at a national level.

Regardless of the inconsistent findings in terms of language background, the current study’s finding again emphasises the importance of developing ELP, particularly for life after university. Language skills are considered vital for employment purposes (Humphreys, 2015). Employability literature confirms the necessity for strong communication skills in English, among various other skills (Gribble & Blackmore, 2012; Humphreys & Gribble, 2013; Kinash & Crane, 2015; Murray & Arkoudis, 2013).

6.1.2.1.4 NON-ATTENDANCE AT ASC

This study also found that students who did not utilise Bond University's ALL unit, Academic Skills Centre or ASC, were 2.89 times more likely to maintain a failing GPA after two semesters compared to those who did. It is often suggested that students who do not engage with skill development initiatives, such as ALL units, for a plethora of reasons, are most frequently the ones who most require it (Murray, 2012; Podorova, 2016). The current study's finding that those who did not engage with support were more likely to fail provides empirical evidence to support this claim.

Because engagement with academic support services is most often non-compulsory (Ransom, 2009), it is unsurprising that students do not engage, even when engagement is incentivised. However, there is a dearth of literature examining the academic performance of students who elect not to engage with support services. Concerning DELNA and the University of Auckland context, it was suggested that most students who avoided participating in the follow up stage went on to maintain low GPAs (Read & von Randow, 2013). Moreover, a longitudinal study by Ashton-Hay and Doncaster (2021), examining the academic performance of 13,239 students over three years at Southern Cross University, found that students who did not attend ALL consultations performed worse academically in terms of both subject completion and GPA. These students were also found to have higher rates of attrition in comparison to those who engaged with support services.

Throughout the literature investigating the services provided by ALL units and their contribution to academic performance, it is commonly stated that providing clarity concerning the relationship is difficult (Ashton-Hay & Doncaster, 2021; Chanock 2007; Stevenson & Kokkinn, 2009). However, studies have demonstrated the important role played by such units in the individual consultations they provide, particularly for developing academic writing skills (Huijser, et al., 2008; O'Mahony et al., 2013; Woodward-Kron, 2007). The finding that students who elect not to engage with the support service were more likely to fail than students who did engage is notable, as it adds weight to the impact of ALL units, such as ASC at the research site.

6.1.2.1.5 BOND UNIVERSITY COLLEGE AND BOND BUSINESS SCHOOL STUDENTS

In regards to enrolment in particular areas of the university and academic performance, students enrolled in a BUC diploma programme (i.e., a sub-bachelor

programme) and students enrolled in BBS were identified as more likely than students from other faculties to fail after two semesters. BUC students were 3.12 times more likely than students from other faculties to be failing, whereas BBS students were 1.74 times more likely to fail.

The finding that BUC students were more than three times as likely to fail after two semesters compared to students from other faculties is somewhat expected based on national data. With the mission to provide students with skills and knowledge they require to transition into their desired bachelor degree programme (Bond University, 2022b), the college offers preparation programmes, such as the Foundation Program, Diploma Preparation Program, and various diploma programs (e.g., Diploma of Arts, Diploma of Business). At a national level, 2020 data demonstrates the attrition rate of all students enrolled in a sub-bachelor programme, such as diplomas, was 31.01% (Department of Education, 2023b). In comparison, the retention rate for all students enrolled in sub-bachelor programmes was 67.56% (Department of Education, 2023b). BUC experienced an attrition rate of 14.75% in 2020, while the retention rate was 64.15%. Meanwhile, the success rate in 2021 was 76.70% for sub-bachelor level students (Department of Education, 2023b). These data indicate the comparably high attrition, yet lower retention and success rates amongst students enrolled in Australian sub-bachelor programmes; this was consistent at the college.

In comparison, BBS students were 1.74 times more likely than students from other faculties to maintain a failing GPA after two semesters. In contrast to the findings for BUC, this result was somewhat surprising, as there is a dearth of literature pointing to business degree students being more inclined to fail in comparison to other degree programmes. It is an important finding but may be contextual to the research site. Anecdotally, it has been suggested that students commencing a business degree, such as a Bachelor of Business, may not be fully aware of what the degree entails. As noted by Tumen et al. (2008, p. 248), “students’ program choice seems to be a critical factor impacting on students’ pathway outcomes, both in terms of completion and non-completion.” Furthermore, it has been suggested that many students commencing business degrees may not be prepared for the amount of mathematics and statistics-based content. This may lead to students being unprepared or ill-equipped to navigate new content and subsequently discontinue their studies. As a consequence, further investigation is necessary to more fully comprehend this finding.

6.1.2.1.6 GENDER (MALES)

The final factor identified to predict academic performance was gender. Specifically, male students were found to be 1.93 times more likely than females to be averaging a fail grade after two semesters in comparison to their female peers. This finding is somewhat consistent with extant literature and reported data.

There have been numerous investigations analysing and comparing educational performance of genders. For instance, in a study of 8,896 undergraduate students at the University of Western Sydney, being male, was associated with low GPA (Grebennikov & Skaines, 2009). Additionally, Smith et al.'s (2012) study determined that males were more at risk of failure in a first-year accounting subject than females. A number of studies, however, have found that if and when gender is a significant factor in determining academic achievement, the margin is minimal. Australian studies have found that there was no performance difference between females and males in science (MacKenzie & Schweitzer, 2001) or medicine for graduate students (Puddey & Mercer, 2014). Harvey and Luckman (2014) also found that demographic factors, such as sex, as well as age, socio-economic status, and regionality, had no significant relationship with attrition within a sample of 1,124 commencing undergraduate Arts students at LaTrobe University. When a relationship has been found between gender and academic performance, it is generally small. Cao and Gabb's (2006) study examining 12,500 commencing students and their attributes at Victoria University in 2002 and then again in 2004 found that males only had a slightly higher rate of attrition. Consistent with this finding was Danilowicz-Gösele et al.'s (2014, p. 21) study which stated that gender and other factors "play only a minor role." Additionally, the New Zealand study by Tumen et al. (2008, p. 245) found that, when controlling for student experiences and achievement, gender, as well as ethnicity and socio-economic status, had "no predictive power".

When examining the attrition, retention and success rate data amongst Australian universities, it appears that the current study's finding regarding males being more inclined to fail after two semesters is a national trend. Data provided by the Tertiary Collection of Student Information (Department of Education, 2023b) demonstrates that, in 2022, male students had a higher rate of attrition (13.35%) in comparison to their female counterparts (11.77%) at bachelor degree level. For "Indeterminate/Intersex/Unspecified" genders, this figure was 18.99%. However, at the sub-bachelor level, including diploma programmes,

attrition was lower for males (27.09%) compared to females (33.67%). The attrition rate for “Indeterminate/Intersex/Unspecified” genders completing a sub-bachelor programme was 31.43%. The rates for retention in 2020 concerning gender were quite similar at the bachelor degree level, yet the rate for males (86.63%) was again lower compared to females (87.37%). “Indeterminate/Intersex/Unspecified” genders had a retention rate of 80.59%. At sub-bachelor level, however, there was increased variance amongst males (71.65%), females (64.78%) and “Indeterminate/Intersex/Unspecified” genders (68.57%). Finally, regarding success rates at bachelor level, in 2021, males had a lower rate of success (83.42%) compared to females (87.14%). “Indeterminate/Intersex/Unspecified” genders had a retention rate of 79.02%. At the sub-bachelor level, males again had a lower rate of success (76.00%) compared to females (77.39%), with “Indeterminate/Intersex/Unspecified” genders having a 74.87% success rate. In sum, the national higher education data points to males generally having a higher attrition rate and lower success rate compared to females, although this was not always the case at the sub-bachelor level. The finding in the current study is in line with national data.

6.1.2.2 PREDICTING HIGH ACHIEVEMENT

The second predictive analysis examined factors predicting students reaching high achievement, defined as a Distinction or 75% average for subjects completed (i.e., obtaining a GPA of 3.00 and above) after two semesters of study. The following characteristics were found to be predictors: obtaining a Satisfactory score on BELA (i.e., 7 or 9), enrolment in a Health Sciences and Medicine (HSM) degree, coming from an English-speaking background, gender (female), and age (older). The odds of maintaining a GPA above 3.0 after two semesters were: 5.16 times higher for students obtaining a Satisfactory score on BELA, 4.63 times greater for students enrolled in HSM, 1.67 times greater for English-speaking background students, 1.45 times greater for females, and 1.03 times greater for older students. The next sections discuss each factor as predictors of high achievement.

6.1.2.2.1 SATISFACTORY WRITING AS MEASURED BY BELA

Students obtaining a Satisfactory score on BELA (i.e., scores of 7 or 9) were found to be 5.16 times higher than those receiving a Below satisfactory score (i.e., scores of 3 or 5) to achieve a distinction average or above. Although limited, the literature investigating PELAs as predictors of academic achievement frequently discusses how those scoring high on PELAs are likely to score high academically, commonly measured by GPA.

Several studies have highlighted the role PELAs play in predicting academic performance. In terms of high academic achievement, Erlam and Botelho de Magalhaes (2021) demonstrated a significant, modest correlation (.33) between DELNA results and GPA. The researchers found that students scoring above the threshold score for DELNA were more inclined to achieve higher GPA outcomes in comparison to students scoring below. Lydster and Brown (2017), which used BELA to examine the extent to which students' performance on the PELA was related to performance on the Major Essay, determined that the balance of grades for students with a Satisfactory BELA score was much higher in comparison to grades for those with a Below satisfactory score. However, there is a paucity of research investigating high performance on PELAs and academic performance; hence, the current study's finding makes a notable contribution.

6.1.2.2.3 HEALTH SCIENCES AND MEDICINE STUDENTS

Concerning faculty, students enrolled in HSM were 4.63 times more likely to achieve a Distinction average compared to students from other faculties/college. This is an important finding, as previous studies have found that students enrolled in natural and physical science degrees (e.g., biology and chemistry) were more likely to discontinue their studies than other cohorts (Scott & Smart, 2005). This may be explained by anecdotal evidence suggesting students enrolled in HSM degrees are inclined to pursue medical degrees, requiring a high GPA for admission. HSM students often leave no stone unturned in order to achieve academic success, and this is similar to Ashton-Hay and Doncaster's (2021) finding that amongst over 13,000 students, approximately half of the students who attended an ALL consultation were studying a degree within the School of Health and Human Sciences faculty. These students attending ALL consultations performed better academically in terms of subject completion and GPA and maintained lower rates of attrition than students who did not utilise the support services. The current study's finding is also consistent with the New Zealand study by Tumen et al. (2008), which found students undertaking human biology and science degrees were, overall, more likely to finish their subjects amongst all third-year students.

6.1.2.2.2 ENGLISH-SPEAKING BACKGROUND STUDENTS

Students from English-speaking backgrounds were 1.67 times likely to average Distinction or above in comparison to EAL/D students. This finding is similar to what has been reported regarding the propensity of English-speaking background students to perform

better academically than their EAL/D student peers (Bougnan, 1993; Gerardi, 1996; Grebennikov & Skaines, 2009; Smith et al., 2012). In contrast, studies have also found no relationship to exist between language background and academic performance (Berman & Cheng, 2010; Wu Pong et al., 1997), as discussed in section 6.1.2.1.3. Some studies and, notably, national data indicate that English-speaking background students have higher rates of attrition (McMillan, 2005) and lower rates of retention and success compared to their EAL/D student counterparts (Department of Education, 2023b). Thus, this finding makes a noteworthy contribution, yet on-going investigations are required.

6.1.2.2.4 GENDER (FEMALES)

Gender was found to be a predictor of high academic achievement after two semesters of study. Specifically, female students were 1.45 times more likely than males to average a distinction or above after two semesters. This finding is consistent with what is routinely found throughout the literature. As discussed in section 6.1.2.1.6, females tend to demonstrate slightly higher academic achievement compared to males (Cao & Gabb 2006; Grebennikov & Skaines, 2009; Mthimunye & Daniels, 2019). Furthermore, Anderton et al. (2016) investigated which variables predicted grades on a health sciences subject. Significant factors included gender, with females performing 2.8% to 7.8% better than their male counterparts over a three-year period (significance was found in two of the three years). Finally, the current finding is consistent with Scott and Smart's (2005) study that found females performed consistently better than males academically and were 1.2 times more likely to complete their degree programme compared to males.

6.1.2.2.5 AGE (OLDER STUDENTS)

The final factor determined to predict high academic achievement was age. Specifically, older students were found to be just 1.03 times more likely than younger students to average a distinction or above after two semesters. This is a minor difference, but indicates that as students get older, they are slightly more inclined to obtain high achievement at university.

When it comes to predicting academic performance, including attrition, age has been found to have only "a minor effect on the likelihood of departure from a program" (Tumen et al., 2008, p. 248) upon controlling for student achievement and factors related to study. Often, studies investigating age as a predictor of academic performance are inconsistent

(Anderton et al., 2016). As noted by McKenzie and Schweitzer (2001), some studies have found that younger students (e.g., recent school leavers) attain higher results and are more inclined to persist with their studies in contrast to more mature students. Specifically, students 19 years old and younger have been found to have higher completion rates compared to other age groups (Cao & Gabb, 2006; Department of Education, Skills and Employment, 2021; Vinke & Jocham, 1993). In contrast, studies have also found the opposite to be true. Several predictive studies have determined that younger students (i.e., younger than 25 years old) are more at risk of academic failure, and older students are more likely to complete their degrees in comparison to recent high school graduates (Mthimunye & Daniels, 2019).

Inconsistent findings have been documented at a national level. Data provided by the Tertiary Collection of Student Information (Department of Education, 2023b) indicates that bachelor degree students in the 19-years and under category maintained a success rate of 86.36% compared to students in the 40-years and over category with 82.77%. This demonstrates that younger students, nationally, had slightly higher success rates at bachelor level. However, these results are flipped at the sub-bachelor level. The success rate for those 19-years and under was 75.12% in comparison to 78.67% for students in the 40-and over category, again highlighting the inconsistency. Consequently, age and level of study requires further investigation to better understand the trends.

In sum, there are contrasting findings throughout the literature and national data. This study found only a slight relationship between age and academic performance. Age in itself is stated not to be a factor affecting the academic achievement of university students; however, it is related to factors, such as prior content knowledge, which may immediately affect success at tertiary level (Vinke & Jocham, 1993). Thus, it is important to be mindful of this when investigating the relationship.

6.2 IMPLICATIONS

There are several implications associated with the current study. The first is the achievement of the study's main aim, to determine whether academic and non-academic factors, such as BELA scores, were able to predict students' academic performance at university. The importance of the first year concerning retention and success is highlighted, as well as the necessity for a whole of university approach to student support. A further implication is that this study has confirmed the ability of PELAs to identify students who

may require additional support. Finally, this thesis lends support regarding the important role played by on-campus ALL units in delivering such support.

6.2.1 ABILITY TO PREDICT ACADEMIC PERFORMANCE

This research was able to determine factors that predicted academic performance of undergraduate students at one Australian university, Bond University. Consequently, profiles of learners likely to fail and, in contrast, maintain high grades were created. Such profiles can be used to assist in the identification of potentially at-risk groups of students. These profiles are considered important outcomes of the current research, as they allow decision makers to allocate resources to ensure potentially at-risk students are given every opportunity to achieve their academic goals. As argued by Tumen et al. (2008, p. 246):

The production of such profiles provides educational institutions with a tool for forecasting the outcomes of students and for the early detection of divergences from a pathway of completion. Armed with such informational tools, institutions could provide positive and appropriate interventions to support students in achieving their goals.

Determining factors that predict academic failure in particular is a notable achievement, as identifying students at risk of failing or discontinuing their studies must be achieved as early as possible. The sooner students and relevant staff are aware of potential challenges, the sooner interventions can occur (Gershenfeld et al., 2015). Hence, identifying students in their first semester who may require additional support provides decision makers with an early alert of students at risk of not graduating.

6.2.2 THE IMPORTANCE OF THE FIRST YEAR

The current study has also emphasised the importance of the first year of study from the lens of a retention and success perspective. As demonstrated throughout the literature on academic performance predictions, students' first year is argued to be the most important in terms of predicting completion (Tinto & Love 1995; Tumen et al., 2008;). Kift (2009, p. 40) maintained that the first-year experience of students is “arguably the most crucial time for engaging students ... not only to persist, but to be successful and independent in their new learning throughout their undergraduate years and for a lifetime of professional practice.” Furthermore, university decision makers can affect policy so that interventions can increase undergraduate students' chances of completion (Tumen et al., 2008). These interventions may

include provision of resources to support units, such as ALL units and Indigenous tutoring services. As argued by Linden et al. (2022, p. 42), “institutional commitment to the student experience in the early stages of university has the greatest potential to exceed student expectations.”

With the introduction of the Higher Education Support Amendment (Job-Ready Graduates and Supporting Regional and Remote Students) Bill 2020 No. 93, 2020 to the Higher Education Support Act 2003 No. 149 (Cth) and, subsequently, the Australian Universities Accord, the ability of universities to predict which students may possibly fail is crucial. A clear focus on equity of access and changes to university funding demonstrate that attrition, retention and success will be high priorities for universities. Thus, efforts to enhance retention and success within higher education seem to have never held such serious repercussions. Part of the strategy to address the Australian Government’s changes is to identify students who are at risk in order to provide support as soon as possible. It is believed the current study contributes to this endeavour.

6.2.3 WHOLE OF UNIVERSITY APPROACH REQUIRED

Given the serious nature of attrition, responding to challenges faced by students requires a whole of university approach. Providing targeted assistance to the students who most require it necessitates “a coordinated, whole of institution support” (Linden et al., 2022, p. 42). This whole of university approach is a view shared by many scholars. Krause (2005, p. 65), for instance, called for “different parts of the institution to work together single-mindedly with the aim of producing a seamless educational experience for students.” A whole of university approach to language and academic literacies development has been demonstrated to be successful in several Australian university contexts (Arkoudis & Doughney, 2014). Goldsmith et al. (2022), for instance, reported the efficacy of a wide-reaching, university-wide academic language enhancement programme at the University of Technology Sydney. The first stage of the programme involved the implementation of an online PELA or OPELA (i.e., a modified version of PAAL), with subsequent follow-up support and language development tutorials (Edwards et al., 2021). The researchers found that almost 4,000 students who participated in the programme not only developed their academic language skills, but were also more confident, made stronger social connections, and were better equipped in navigating academia, demonstrating an exemplar of good practice (TEQSA, 2020).

An often-quoted statement concerning admission and completion at university is, “access is meaningless without success” (Murray, 2016, p. 130). In line with this sentiment, it is imperative that universities proffer support to all students who require it (Lobo, 2012). The first step in achieving this is to ensure that all students are aware of the on-campus support available. It is believed that the current study makes an important contribution by presenting a validity argument for the effective use of a PELA in achieving awareness of academic support provided on campus by an ALL unit.

6.2.4 PELAS CAN IDENTIFY AT-RISK STUDENTS

A further implication of the current study is the evidence found to support the use of a PELA in effectively identifying students potentially at risk of academic failure. This study has demonstrated that students whose writing is assessed as Below satisfactory via BELA are considerably more likely to fail the Major Essay and perform worse on the subject’s overall grade, as well as maintain lower GPAs in comparison to students obtaining Satisfactory scores. Consequently, it is necessary that this group of students is afforded as much support as possible. An additional benefit of implementing the BELA process at Bond University has been the increased awareness amongst stakeholders, including students, academic staff, professional staff, and decision-makers, of potential challenges early in students’ degrees, allowing targeted support to occur as early as possible. Given the governmental changes higher education policy, it is critical that students experiencing difficulties are noticed. Once a student’s writing has been identified as Below satisfactory, the chain of support can begin with a consultation at ASC. Detailed feedback can be discussed with a Learning Advisor and any issues with academic writing can be identified and suitable resources suggested. Furthermore, the coordination between academic staff who teach on Core 1: Critical Thinking and Communication and Learning Advisors at ASC assist in achieving the scaffolding of first year student learning, as recommended by Kift et al. (2010). This collaboration has resulted in beneficial support of students and contributed to a positive learning experience in students’ first semester of study. As opined by Erlam and Botelho de Magalhaes (2021, p. 35), “a PELA programme must meet the needs of the institution it serves.” It is argued that this has been achieved.

6.2.5 THE ROLE PLAYED BY ALL UNITS

A final implication of this research is that it highlights the important role played by support units, such as ASC at the University. Although studies are emerging that provide

valuable insight into the important role played by academic language and learning units (e.g., Edwards et al., 2021; Knoch, 2012), it has proven difficult to measure impact. As noted by Ashton-Hay and Doncaster (2021, p. 104), “few studies have been able to successfully establish a link between ALL work, student success and retention.” The current research has provided substantial insight into ASC’s impact at the research site. Although many challenges exist, such as equitable and sustainable provisions of support (Green et al., 2017; Murray, 2022), evidence has been presented for the impact of an ALL unit in terms of student success. This evidence can inform decision makers who set university policy to better support the student experience. This is timely as universities lack the required resources to support students in overcoming language and academic literacies challenges to achieve success (Barthel, 2023).

6.3 LIMITATIONS AND FUTURE RESEARCH

Despite the contributions of this research, several limitations should be noted. The first is the use of the term, success. As introduced in Chapter 2 of this thesis, success is often used to define academic completion. York et al. (2015, p. 1) described the concept as “one of the most widely used constructs in educational research and assessment within higher education.” However, the researchers determined its definition is complex, broad and, at times, misused, with academic achievement being just one factor in their theoretical definition. In a chapter written by three university students titled, “Success at university: The student perspective”, Hannon et al. (2017) explained how one student’s definition of academic success is dependent on the individual’s background, field of study, previous achievements, and personal standards, amongst other factors. A common theme from Hannon et al.’s (2017) work was that happiness played a substantial role in students’ conception of success, as well as confidence and mental health. Additionally, a study undertaken at the research site enquiring into definitions of success amongst Indigenous students found that participants defined success as not only maintaining a high GPA, but about future endeavours, for instance, establishing a pathway to employment or further studies (Lydster & Murray, 2018). The students in the study commented that success at university entailed breaking through cultural stereotypes and setting examples for others. Indicative quotes included: “success is overcoming that ... ‘you’re Indigenous, you don’t go to uni’, sort of stereotype” and “I think for me it’s just like telling other kids and my cousins and everyone at home that they can do it too. That for me is success” (Lydster & Murray, 2018, p. 113). Success at university has multiple meanings and each stakeholder group may view success

from different lenses, presenting an opportunity to further investigate stakeholders' concepts of success.

Furthermore, as highlighted throughout the literature review, there is an abundance of factors that play a role in academic performance. This study was positioned from an applied linguistic perspective, combining the results of a language assessment and other demographic factors, as well as academic performance variables. Future studies may seek to broaden the scope and include factors such as socio-economic status, a common focus of predictive studies. Students who represent low socioeconomic status may be faced with additional challenges to achieving success in their studies (Grebennikov & Skaines, 2009; Smith et al., 2012). A meta-analytic review of literature between 1990 and 2000, including a large sample of 101,157 students, by Sirin (2005) demonstrated that socioeconomic status impacted students' academic performance both directly in the form of learning resources at home and indirectly through social capital. However, there is much debate and inconsistent findings in the literature, with several studies finding socio-economic status was inadequate as a standalone category to predict academic failure at university (Carpenter et al., 1998; Harvey & Luckman, 2014; Walker-Gibbs et al., 2019). An Australian study by Krause et al. (2005, p. 68) maintained that "the available data on students from lower socio-economic backgrounds shows that while access rates are lower, students once enrolled have broadly comparable rates of success, retention and completion." The research site for this current thesis, Bond University, has been found to have an atypically low share of students from low socioeconomic backgrounds (Leigh, 2021), yet student demographics are changing, and there exists an opportunity for future research to investigate socioeconomic status' role in academic achievement.

There may also be an opportunity for collaboration with experts in the field of educational psychology. Psychological variables, such as personality factors, self-efficacy and motivation, are often cited as having an association with academic performance (Graham, 1987; Ho & Spinks, 1985; Malczewska-Webb, 2016; Mthimunye & Daniels, 2019). For example, an Australian study by Burton and Dowling (2005) found personality traits to be predictors of academic performance (i.e., GPA) amongst on campus, engineering students. Notably, conscientiousness was found to be a positive predictor, whilst extraversion was found to be a negative predictor of academic performance. Other psychological factors that have been considered include self-efficacy and motivation (Burton & Dowling, 2005;

McKenzie & Schweitzer, 2001; Mthimunya & Daniels, 2019; Reynolds & Weigand, 2010). Throughout the educational literature, self-efficacy is seen as imperative for success at university (Fenton-Smith et al., 2017; Pajares, 1996; Rochecouste et al., 2011). A systematic review of 7,167 studies comprising 241 data sets by Richardson et al. (2012) found performance self-efficacy had the strongest correlation of the 50 measures investigated, with academic self-efficacy found to have a moderate sized correlation with GPA. Moreover, motivation and effort to study is argued to be a factor determining academic performance (Burton & Dowling, 2005; Gue & Holdaway 1973; Reynolds & Weigand, 2010). Matsumoto (2011), whose study was conducted at the University, found that international EAL/D students' motivation to learn was linked to perception of their instructors. It was beyond the scope of the current research to investigate such psychological factors and their relationship with academic performance; however, there may be an opportunity for future research to include such factors.

Additional research opportunities include gathering further evidence to support the validity argument of BELA at the research site. Investigations into DELNA, for example, highlight the importance of ongoing validation of PELAs, as opposed to “one off” efforts (e.g., Erlam & Botelho de Magalhaes, 2021; Read, 2015a). The current investigation, for instance, examined the inter-rater reliability of raters using Classical Test Theory. It did not examine intra-rater reliability or the internal consistency of a single rater (Cho, 1999). Cho (1999, p. 1) argued that intra-rater reliability is “as significant as inter-rater reliability, since if the former is not secure, neither is the latter.” Due to resource constraints, it was not possible to engage academic staff in further rounds of rating BELA essays. Additionally, future research could employ sophisticated methods of analysis, such as Rasch modelling (Fan et al., 2019; Fan & Knoch, 2019), to investigate reliability, including intra-rater reliability, and assess the efficacy of online rater training.

An additional limitation regards the lack of reliability concerning aspects of the self-created questionnaires to gain stakeholders' insight on the validity of BELA. Although moderate to good reliability was determined for the majority of inferences developed in the three questionnaires, low reliability was determined for the Generalisability items of the student questionnaire, and a negative reliability was found for the Decisions items of the academic staff questionnaire. This limitation was particularly evident for the latter questionnaire. It appeared that academic staff either agreed that feedback was sufficient but

did not believe it was accessible or vice versa, that feedback was accessible but not sufficient. Again, this points to confusion regarding the provision of feedback via BELA. Further research, for instance interviews and/or focus groups with stakeholders, is required to pinpoint the reason for a perceived lack of feedback from the PELA.

As part of an ongoing investigation into BELA's validity, future research may also investigate the feasibility and impact of broadening the target group of students post-BELA completion. This suggestion is in response to two concerns: the perceived lack of feedback for students once they have completed BELA, most likely amongst students scoring 7 or 9, and lack of consistency amongst raters. This would involve strongly recommending students scoring 7 on BELA to also attend Academic Skills Centre consultations. As the Satisfactory category includes students scoring 7 of a possible score of 9, it indicates that one area of the student's academic essay writing has been identified as lacking. Rather than necessitating only students scoring 3 or 5 on BELA to attend consultations at ASC, it may be beneficial to also encourage students who scored 7 to receive detailed feedback Learning Advisors. This may assist in minimising any issues associated with raters incorrectly rating students 7 (i.e., Satisfactory) instead of 5 (i.e., Below satisfactory).

6.4 CONCLUSION OF THESIS

In conclusion, this thesis achieved the aim of determining factors able to predict performance amongst undergraduate students. These factors included: Indigeneity, scoring 3 or 5 on BELA (Below satisfactory), enrolment in a Bond University College or Bond Business School programme, not utilising Academic Skills Centre consultations, using English as an additional language or dialect, and being male. Regarding high academic achievement, the characteristics of students likely to attain a Distinction or above included: scoring 7 or 9 on BELA (Satisfactory), enrolment in an Health Sciences and Medicine degree, coming from an English-speaking background, being female, and being older. Bond University is now equipped with profiles of learners and their predicted academic performance after two semesters of study. These profiles allow policymakers to make informed decisions regarding allocation of appropriate resources. In particular, it allows those most at risk of academic failure to be targeted, supported, and, hopefully, equipped with the necessary knowledge, skills, and confidence to succeed in academia.

As the main research tool used in the study was the Bond English Language Assessment (BELA), it was imperative present a validity argument for the PELA. Overall, a

satisfactory validity argument was presented for BELA. Each inference was supported, and BELA and its support processes are well placed within a university policy context that promotes its reach to the majority of undergraduate students. As a consequence, whole cohorts of undergraduate students at the University can be screened based on their academic essay writing skills. Subsequently, those who may benefit from additional assistance are provided with effective and appropriate on-campus support. The current study is notable, as it combines the most detailed presentation of validity evidence, using an established framework, to present an argument for the implementation of BELA and the use of its scores.

Nevertheless, concerns were identified regarding the PELA. Most critical were the findings in terms of the ability of raters to consistently apply the rubric. It was, therefore, recommended that an online rater training module be created for both new and continuing raters to refer to. This could be delivered whilst also monitoring the ratings of academic staff in order to enhance consistency. Additional concerns regarded academic integrity and, with the development of generative AI software such as Chat GPT, these concerns will increase. A final concern was feedback, or a lack thereof. On one hand, it is not the intention of BELA to act primarily as a diagnostic measure, and therefore provide feedback to each test taker. On the other hand, it seems like a missed opportunity to not promote ways learners can develop their academic writing skills whilst completing their studies. At a minimum, general advice should be provided for all students. Overall, considering the low-stakes nature of BELA, an adequate validity argument has been presented.

In conclusion, it is important to note that the results of this study may not be generalisable to other university contexts. However, it may encourage researchers to investigate whether similar factors play a role. It is predicted that such endeavours will achieve worthwhile aims of reducing attrition, whilst enhancing retention and success. The reward is very much worthy of the effort.

REFERENCES

- Adams, T., Banks, M., Davis, D., & Dickson, J. (2010). *The Hobsons retention project: Context and factor analysis* report. <https://silo.tips/download/the-hobsons-retention-project-context-and-factor-analysis-report>
- Affendey, L. S., Paris, I. H. M., Mustapha, N., Sulaiman, N., & Muda, Z. (2010). Ranking of influencing factors in predicting students' academic performance. *Information Technology Journal*, 9(4), 832-837. 10.3923/itj.2010.832.837
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge University Press.
- Allen, P., Bennett, K., & Heritage, B. (2019). *SPSS statistics: A practical guide* (4th Ed.). Cengage.
- American Educational Research Association, The American Psychological Association, & The National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Anderson, P. J., Yip, S. Y., & Diamond, Z. M. (2022). Universities Australia 2017–2020 Indigenous Strategy: A meta-synthesis of the issues and challenges. *Higher Education Research & Development*. <https://doi.org/10.1080/07294360.2022.2123899>
- Anderton, R. S., Evans, T., & Chivers, P. T. (2016). Predicting academic success of health science students for first year anatomy and physiology. *International Journal of Higher Education*, 5(1), 250-260. <https://doi.org/10.5430/ijhe.v5n1p250>
- Arkoudis, S., & Doughney, L. (2014). *Good practice report: English language proficiency*. Office for Learning and Teaching.
- Ashton-Hay, S., & Doncaster, N. (2021). Student success and retention: What's academic skills got to do with it? *Journal of Academic Language & Learning*, 15(1), 102-116. <https://journal.aall.org.au/index.php/jall/article/view/727/435435521>
- Asif, R., Merceron, A., & Pathan, M. K. (2015). Predicting student academic performance at degree level: A case study. *I.J. Intelligent Systems and Applications*, 1, 49-61. <https://doi.org/10.5815/ijisa.2015.01.05>
- Avdi, E. (2012). IELTS as a predictor of academic achievement in a Master's Program. *English Australia Journal*, 26(2), 42-49. <https://search.informit.com.au/fullText;dn=361619889813753;res=IELHSS>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34.
https://www.tandfonline.com/doi/pdf/10.1207/s15434311laq0201_1
- Bachman, L. F., & Palmer, A. S (1996). *Language testing in practice*. Oxford University Press.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Baird, C., & Dooley, P. (2017). Using images to facilitate writing for skills assessment: A visual PELA. *The Australian Journal of Indigenous Education*, 46(2), 160-172.
<https://doi.org/10.1017/jie.2016.32>
- Barrett-Lennard, S., Dunworth, K., & Harris, A. (2011). The Good Practice Principles: Silver bullet or starter gun? *Journal of Academic Language & Learning*, 5(2), A99-A106.
- Barthel, A. (2017). *Do you PELA or not?* <https://aall.org.au/wp-content/uploads/2019/09/PELA-Resources-WA-ALL.pdf>
- Barthel, A. (2023). Universities lack resources in academic language and learning. *Campus Morning Mail*. <https://campusmorningmail.com.au/news/universities-lack-resources-in-academic-language-and-learning/>
- Bayliss, A., & Ingram, D. (2006). *IELTS as a Predictor of Academic Language Performance*. Australian International Education Conference.
[http://aiec.idp.com/uploads/pdf/BaylissIngram%20\(Paper\)%20Wed%201630%20MR5.pdf](http://aiec.idp.com/uploads/pdf/BaylissIngram%20(Paper)%20Wed%201630%20MR5.pdf)
- Bellingham, L. (1993). The relationship of language proficiency to academic success for international students. *New Zealand Journal of Educational Studies*, 30(2), 229-232.
- Berman, R., & Cheng, L. (2010). English academic language skills: Perceived difficulties by undergraduate and graduate students, and their academic achievement. *Canadian Journal of Applied Linguistics*, 4(1), 25-40.
<https://journals.lib.unb.ca/index.php/CJAL/article/view/19830>
- Bernat, E., & Gvozdenko, I. (2005). Beliefs about Language Learning: Current Knowledge, Pedagogical Implications, and New Research Directions. *TESL-EJ*, 9(1), 1-21.
<https://files.eric.ed.gov/fulltext/EJ1065832.pdf>
- Birrell, B. (2006). Implications of low English standards among overseas students at Australian universities. *People and Place*, 14(4), 53-64.
<https://search.informit.com.au/documentSummary;dn=332480460154352;res=IELAPA>

- Birrell, B., Hawthorne, L., & Richardson, S. (2006). *Evaluation of the general skilled migration categories*. Canberra, ACT: Commonwealth of Australia.
http://www.flinders.edu.au/sabs/nils-files/reports/GSM_2006_Full_report.pdf
- Bonanno, H. (2002). *Standing the test of time: Revisiting a first year diagnostic procedure* Changing agendas "Te ao hurihuri" (Proceedings of the Sixth Pacific Rim Conference on First Year in Higher Education), Brisbane.
http://www.fyhe.qut.edu.au/past_papers/papers02.htm
- Bonanno, H., & Jones, J. (2007). *The MASUS procedure: Measuring the academic skills of university students a diagnostic assessment*.
<https://www.sydney.edu.au/content/dam/corporate/documents/about-us/partnership/masus-procedure.pdf>
- Bond University. (2022a). *Beyond Bond*. <https://bond.edu.au/current-students/opportunities/beyond-bond>
- Bond University. (2022b). *Bond University College*. <https://bond.edu.au/our-university/our-faculties-and-academic-units/bond-university-college>
- Bond University Student Administration. (n.d.). *Grade point and percentage averages*.
https://bond.edu.au/files/440/GPA_Calc_guide.pdf
- Botha, W. (2022). Academic literacy skills and the challenge for Australian higher education. *Educational Studies*. <https://doi.org/https://doi.org/10.1080/03055698.2022.2130178>
- Boughan, K. (1993). *Nursing student performance 1986-1993: Preliminary findings*.
<https://files.eric.ed.gov/fulltext/ED371805.pdf>
- Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). *Review of Australian higher education: Final report*. <https://www.voced.edu.au/content/ngv:32134>
- Bridgeman, B., Cho, Y., & DiPietro, S. (2016). Predicting grades from an English language assessment: The importance of peeling the onion. *Language Testing*, 33(3), 307-318.
<https://doi.org/10.1177/0265532215583066>
- Bright, C., & von Randow, J. (2004). *Tracking language test consequences: The student perspective*. 18th IDP Australian International Education Conference, Sydney.
- Bright, C., & von Randow, J. (2008). Missed opportunities: EAL students reflect on their first-year language experiences at an English-medium university.
https://unistars.org/past_papers/papers08/FYHE2008/content/pdfs/5d.pdf
- British Council. (2017). *IELTS numbers rise to three million a year*.
<https://www.britishcouncil.org/contact/press/ielts-numbers-rise-three-million-year>
- Brown, A., & Jaquith, P. (2011). the development and validation of an online rater training

- and marking system: Promises and pitfalls. In B. O'Sullivan (Ed.), *Language testing: Theories and practices* (pp. 244-261). Palgrave MacMillan.
- Brunsdon, V., Davies, M., Shevlin, M., & Bracken, M. (2000). Why do HE students drop out? A test of Tinto's model. *Journal of Further and Higher Education*, 24(3), 301-310. <https://doi.org/10.1080/030987700750022244>
- Burton, L. J., & Dowling, D. G. (2005). *In search of the key factors that influence student success at university*. Higher education in a changing world, Sydney, Australia.
- Buxton, L. (2017). Ditching deficit thinking: Changing to a culture of high expectations. *Issues in Educational Research*, 27(2), 198-214. <http://www.iier.org.au/iier27/buxton.pdf>
- Cao, Z., & Gabb, R. (2006). *Student attrition at a new generation university* The Australian Association for Research in Education, Adelaide. <https://www.aare.edu.au/publications-database.php/4983/student-attrition-at-a-new-generation-university-in-australia>
- Carpenter, P. G., Hayden, M., & Long, M. (1998). Social and economic influences on graduation rates from higher education in Australia. *Higher Education*, 35(4), 399-422. <https://www.jstor.org/stable/3448267>
- Chanock, K. (2007). Valuing individual consultations as input into other modes of teaching. *Journal of Academic Language & Learning*, 1(1), A1-A9. <https://journal.aall.org.au/index.php/jall/article/view/1>
- Chapelle, C. (2011). Validation in language assessment. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (Vol. 2, pp. 717-730). Routledge.
- Chapelle, C. A. (1999). Validity in assessment. *Annual Review of Applied Linguistics*, 19, 254-272. <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/1B08C26848F89D0D2E148129AA788918/S0267190599190135a.pdf/div-class-title-validity-in-language-assessment-div.pdf>
- Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a Validity Argument for the Test of English as a Foreign Language* (pp. 319-352). Routledge. <https://doi.org/10.4324/9780203937891>
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). Test score interpretation and use. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a Validity Argument for the Test of English as a Foreign Language* (pp. 1-26). Routledge. <https://doi.org/10.4324/9780203937891>

- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29(1), 3-13. <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/j.1745-3992.2009.00165.x>
- Chapelle, C. A. (2012). Conceptions of validity. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing* (pp. 21-33). Routledge.
- Chapelle, C. A., & Lee, H. (2021). Understanding Argument-Based Validity in Language Testing. In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies in validation research* (pp. 19-44). Cambridge University Press.
- Chapelle, C. A., & Voss, E. (2021a). Introduction to Validity Argument in Language Testing and Assessment. In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of validation research* (pp. 1-15). Cambridge University Press.
- Chapelle, C. A., & Voss, E. (2021b). Validation research and its limits. In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of validation research* (pp. 325-344). Cambridge University Press.
- Cho, D. (1999). A study on ESL writing assessment: Intra-rater reliability of ESL compositions. *Melbourne Papers in Language Testing*, 8(1), 1-24. https://arts.unimelb.edu.au/_data/assets/pdf_file/0003/3518175/08_1_1_Cho.pdf
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Commonwealth Register of Institutions and Courses for Overseas Students. (n.d.). *Bond University Limited*. Australian Government Department of Education. <https://cricos.education.gov.au/Institution/InstitutionDetails.aspx?ProviderCode=00017B>
- Cotton, F., & Conrow, F. (1998). *An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania*. International English Language Testing System (IELTS) Research Reports. https://www.ielts.org/-/media/research-reports/ielts_rr_volume01_report4.ashx
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. <https://doi.org/https://doi.org/10.1080/14703297.2023.2190148>
- Craven, E. (2012). *The quest for IELTS Band 7.0: Investigating English language proficiency development of international students at an Australian university* (IELTS Research Reports). <https://search.informit.com.au/fullText;dn=160297895959891;res=IELHSS>
- Creswell, J. W. (2008). *Designing and conducting mixed methods research*. Sage.

- Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications
- Cronbach, L. J. (1980). Validity on parole: How can we go straight? *New directions for testing and measurement: Measuring achievement over a decade*, 5, 99-108.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). Harper & Row.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://oce.ovid.com/article/00006823-195507000-00001/PDF>
- Curnow, T. J., & Liddicoat, A. J. (2008). Assessment as learning: Engaging students in academic literacy in their first semester. In A. Duff, D. Quinn, M. Green, K. Andre, T. Ferris & S. Copland (Eds.), *Proceedings of the ATN assessment conference 2008: Engaging students in assessment*. <http://www.ojs.unisa.edu.au/index.php/atna/issue/view/ISBN%20978-0-646-504421/showToc>.
- Daller, M. H., & Phelan, D. (2013). Predicting international student study success. *Applied Linguistics Review*, 4(1), 173-193. <https://doi.org/10.1515/applirev-2003-0008>
- Daller, M., & Wang, Y. (2017). Predicting study success of international students. *Applied Linguistics Review*, 8(4), 355-374. <https://doi.org/https://doi.org/10.1515/applirev-2016-2013>
- Dang, C. N., & Dang, T. N. Y. The predictive validity of the IELTS test and contribution of IELTS preparation courses to international students' subsequent academic study: Insights from Vietnamese international students in the UK. *RELC Journal*, 54(1), 84-98. <https://doi.org/https://doi.org/10.1177/0033688220985533>
- Danilowicz-Gösele, K., Meya, J., Schwager, R., & Suntheim, K. (2014). Determinants of students' success at university. *Discussion Papers, Centre for European Governance and Economic Development*, 214, 1-31. <https://www.econstor.eu/bitstream/10419/102397/1/797624414.pdf>
- Davies, A., & Elder, C. (2005). Validity and validation in language testing. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 795-814). Lawrence Erlbaum Associates.
- Department of Education. (2019). *2018 list of higher education institutions*. Australian Government. <https://www.education.gov.au/higher-education->

- statistics/resources/2018-list-higher-education-institutions
- Department of Education. (2022). *Selected higher education statistics – 2021 student data*.
<https://www.education.gov.au/higher-education-statistics/student-data/selected-higher-education-statistics-2021-student-data>
- Department of Education (2023a). *Australian Universities Accord*.
<https://www.education.gov.au/australian-universities-accord>
- Department of Education. (2023b). *Selected higher education statistics – 2021 student data*.
<https://www.education.gov.au/higher-education-statistics/student-data/selected-higher-education-statistics-2021-student-data>
- Department of Education and Training. (2018). *Final report: Improving retention, completion and success in higher education*. <https://www.education.gov.au/higher-education-statistics/resources/higher-education-standards-panel-final-report-improving-retention-completion-and-success-higher>
- Department of Education, Employment and Workplace Relations. (2009). *Transforming Australia's higher education system*. <https://www.voced.edu.au/content/ngv:14895>
- Department of Education, Skills and Employment. (2021). *Completion rates - cohort analyses*. <https://www.dese.gov.au/higher-education-statistics/student-data/completion-rates-cohort-analyses>
- DiGregorio, K.D., Farrington, S., & Page, S. (2000). Listening to our students: Understanding the factors that affect Aboriginal and Torres Strait Islander students' academic success. *Higher Education Research & Development*, 19(3), 297–309.
doi:10.1080/758484344.
- Dobinson, T., & Dunworth, K. (2009). Going it alone: Issues of transparency and accountability in the use of an in-house university test of English. *English Australia Journal*, 25(2), 41-48.
<https://search.informit.org/doi/pdf/10.3316/informit.466989342278874>
- Dobson, I. R., & Skuja, E. (2005). Secondary schooling, tertiary entry ranks and university performance. *People and Place*, 13(1), 53-62.
<https://search.informit.org/doi/pdf/10.3316/ielapa.374963634623048>
- Dobson, I., & Sharma, R. (1993). Student progress: A study of the experience in Victorian tertiary institutions. *Journal of Tertiary Education Administration*, 15(2), 203-211.
<https://doi.org/https://doi.org/10.1080/1036970930150206>
- Dodd, T. (2020). 'Debate stifled by Chinese university students'. *The Australian*.
<https://www.theaustralian.com.au/higher-education/debate-stifled-by-chinese->

- university-students/news-story/95420e14871f1e0935fc773fb64a69be
- Donohue, J. P., & Erling, E. J. (2012). Investigating the relationship between the use of English for academic purposes and academic attainment. *Journal of English for Academic Purposes, 11*, 210-219. <https://doi.org/10.1016/j.jeap.2012.04.003>
- Dooley, P. (1999). An investigation into the predictive validity of the IELTS Test as an indicator of future academic success. *Proceedings of the 8th Annual Teaching Learning Forum*, University of Western Australia.
- Dooley, P., & Oliver, R. (2002). An investigation into the predictive validity of the IELTS Test as an indicator of future academic success. *Prospect, 17*(1), 36-54. <https://espace.curtin.edu.au/handle/20.500.11937/77607>
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, Administration, and Processing* (2nd ed.). Taylor & Francis Group.
- Dunworth, K. (2009). An investigation into post-entry English language assessment in Australian universities. *Journal of Academic Language & Learning, 3*(1), A1-A13. <https://journal.aall.org.au/index.php/jall/article/view/67/57>
- Dunworth, K. (2013a). *Degrees of proficiency: Building a strategic approach to university students' English language assessment and development*. Office for Learning and Teaching. https://ltr.edu.au/resources/ID11-2048_Dunworth_Report_2013.pdf
- Dunworth, K. (2013b). *In-course student English Language Development*. Five years on: English language competence of international students (pp. 52-74). International Education Association of Australia. <https://www.ieaa.org.au/documents/item/54>
- Dursun, A., & Li, Z. (2021). A systematic review of argument-based validation studies in the field of language testing (2000–2018). In C. A. Chapelle & E. Voss (Eds.), *Validity argument in language testing: Case studies of validation research* (pp. 45-70). Cambridge University Press.
- Edwards, E., Goldsmith, R., Havery, C., & James, N. (2021). An institution-wide strategy for ongoing, embedded academic language development: Design, implementation and analysis. *Journal of Academic Language & Learning, 15*(1), 53-71. <https://journal.aall.org.au/index.php/jall/article/view/745>
- Elder, C. (1993). Language proficiency as a predictor of performance in teacher education. *Melbourne Papers in Language Testing, 2*(1), 72-95.
- Elder, C. (2017). Language assessment in higher education. In Shohamy, E., Or, I., May, S. (Eds.), *Language Testing and Assessment, Encyclopedia of Language and Education*

- (pp. 271-286). Springer. https://doi.org/10.1007/978-3-319-02326-7_35-1
- Elder, C., Barkhuizen, G., Knoch, U., & von Randow, J. (2007). Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing*, 24(1), 37-64. <https://doi.org/10.1177/0265532207071511>
- Elder, C., Bright, C., & Bennett, S. (2007). The role of language proficiency in academic success: Perspectives from a New Zealand university. *Melbourne Papers in Language Testing*, 12(1), 24-58. http://wayback.archive-it.org/1148/20130404065745/http://ltrc.unimelb.edu.au/mplt/papers/12_1_2_Elder.pdf
- Elder, C., & Erlam, R. (2001). *Development and validation of the Diagnostic English Language Needs Assessment (DELNA): Final report*. University of Auckland, Department of Applied Language Studies and Linguistics. <https://www.auckland.ac.nz/en/students/student-support/academic-support/delna/bibliography.html>
- Elder, C., Knoch, U., Barkhuizen, G., & von Randow, J. (2005). Individual feedback to enhance rater training: Does it work? *Language Assessment Quarterly: An International Journal*, 2(3), 175-196. https://doi.org/https://doi.org/10.1207/s15434311laq0203_1
- Elder, C., Knoch, U., & Zhang, R. (2009). Diagnosing the support needs of second language writers: Does the time allowance matter? *TESOL Quarterly*, 43(2), 351-360. <https://www.jstor.org/stable/pdf/27785015.pdf>
- Elder, C., McNamara, T., & Congdon, P. (2003). Rasch techniques for detecting bias in performance assessments: An example comparing the performance of native and non-native speakers on a test of academic English. *Journal of Applied Measurement*, 4(2), 181-197.
- Elder, C., & Read, J. (2015). Post-entry language assessment in Australia. In J. Read (Ed.), *Assessing English Proficiency for University Study* (pp. 25-46). Palgrave MacMillan.
- Elder, C., & von Randow, J. (2008). Exploring the utility of a web-based English language screening tool. *Language Assessment Quarterly*, 5(3), 173-194. <https://doi.org/10.1080/15434300802229334>
- Erlam, R. & Botelho de Magalhaes, M. (2021). Post-Entry English Language Assessment at the University of Auckland: Ongoing validation of DELNA. *New Zealand Studies in Applied Linguistics*, 21(1), 32-49. <https://www.alanz.org.nz/journal/post-entry-english-language-assessment-at-the-university-of-auckland-ongoing-validation-of->

delna/

- Erling, E. J., & Richardson, J. T. E. (2010). Measuring the academic skills of university students: Evaluation of a diagnostic procedure. *Assessing Writing, 15*, 177-193. <https://doi.org/10.1016/j.asw.2010.08.002>
- Evans, M., & Farley, A. (1998). *Institutional characteristics and the relationship between students' first-year university and final-year secondary school academic performance*. Monash University, Department of Econometrics and Business Statistics
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.439.3676&rep=rep1&type=pdf>
- Fan, J., & Knoch, U. (2019). Fairness in language assessment: What can the Rasch model offer? *Papers in Language Testing and Assessment, 8*(2), 117-142.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0009/3241629/8_2_S5_Fan_and_Knoch.pdf
- Fan, J., Knoch, U., & Bond, T. (2019). Application of Rasch measurement theory in language assessment: Using measurement to enhance language assessment research and practice. *Papers in Language Testing and Assessment, 8*(2), iii-ix.
https://www.altanz.org/uploads/5/9/0/8/5908292/1._plta_8_2__editorial.pdf
- Feast, V. (2002). The Impact of IELTS Scores on Performance at University. *International Education Journal, 3*(4), 70-85.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.457.4736&rep=rep1&type=pdf>
- Fenton-Smith, B., & Humphreys, P. (2015). Language specialists' views on academic language and learning support mechanisms for EAL postgraduate coursework students: The case for adjunct tutorials. *Journal of English for Academic Purposes, 20*, 40-55. <https://doi.org/https://doi.org/10.1016/j.jeap.2015.05.001>
- Fenton-Smith, B., Humphreys, P., Walkinshaw, I., Michael, R., & Lobo, A. (2017). Implementing a university-wide credit-bearing English language enhancement program: Issues emerging from practice. *Studies in Higher Education, 42*(3), 463-479. <https://doi.org/https://doi.org/10.1080/03075079.2015.1052736>
- Ferguson, G., & White, E. (1998). A small-scale study of predictive validity. *Melbourne Papers in Language Testing, 7*(2), 15-63. http://wayback.archive-it.org/1148/20130404070127/http://ltrc.unimelb.edu.au/mplt/papers/07_2_2_Ferguson&White.pdf

- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. <https://doi.org/10.1037/h0031619>
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions* (3rd ed.), Wiley, Hoboken. <https://doi.org/10.1002/0471445428>
- Fox, J. (2004). Test decisions over time: Tracking validity. *Language Testing*, 21(4), 437-465. <https://doi.org/10.1191/0265532204lt292oa>
- Fox, J., Haggerty, J., & Artemeva, N. (2016). Mitigating risk: The impact of a diagnostic assessment procedure on the first-year experience in Engineering. In J. Read (Ed.), *Post-admission language assessment of university students* (pp. 43-66). Springer.
- Fox, J., von Randow, J., & Volkov, A. (2016). Identifying students at-risk through post-entry diagnostic assessment: An Australasian approach takes root in a Canadian university. In V. Aryadoust & J. Fox (Eds.), *Trends in language assessment research and practice* (pp. 266-285). Cambridge Scholars Publishing.
- Fulcher, G. (2010). *Practical Language Testing*. Hodder Education.
- Fulcher, G. (2015). *Re-examining language testing: A philosophical and social inquiry*. Routledge.
- Fulcher, G., & Davidson, F. (2012). Introduction. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 1-17). Routledge
- Gerardi, S. (1996). *The effects of English as a second language on college academic outcomes*. Educational Resources Information Centre (ERIC). <https://files.eric.ed.gov/fulltext/ED398946.pdf>
- Gershenfeld, S., Hood, D. W., & Zhan, M. (2015). The role of first-semester GPA in predicting graduation rates of underrepresented students. *Journal of College Student Retention: Research, Theory & Practice*, 17(4), 469-488. <https://doi.org/10.1177/1521025115579251>
- Ginther, A., & Yan, X. (2018). Interpreting the relationships between TOEFL iBT scores and GPA: Language proficiency, policy, and profiles. *Language Testing*, 35(2), 161-322. <https://doi.org/https://doi.org/10.1177/0265532217704010>
- Goldsmith, R., Havery, C., Edwards, E., James, N., Murphy, A., Mort, P., Nixon, D., O'Donoghue, G., Yang, J. S., & Yeo, J. (2022). A multi-faceted evaluation of the impact on students of an Australian university-wide academic language development program. *Journal of English for Academic Purposes*, 60. 1-15. <https://doi.org/https://doi.org/10.1016/j.jeap.2022.101192>
- Graham, J. G. (1987). English language proficiency and the prediction of academic success.

- TESOL Quarterly*, 21(3), 505-521.
<https://doi.org/https://www.jstor.org/stable/pdf/3586500.pdf>
- Gravatt, B., Richards, J. C., & Lewis, M. (1997). *Language needs in tertiary studies: ESL students at the University of Auckland*. Institute of Language Teaching and Learning Occasional Papers, No. 10. University of Auckland.
- Grebennikov, L., & Shah, M. (2012). Investigating attrition trends in order to improve student retention. *Quality Assurance in Education*, 20(3), 223-236.
<https://doi.org/10.1108/09684881211240295>
- Grebennikov, L., & Skaines, I. (2009). University of Western Sydney students at risk: Profile and opportunities for change. *Journal of Institutional Research*, 14(1), 58-70.
- Green, T., Dymock, J., & Floyd, C. (2017). Academic literacy support: Teaching along the continuum. In L. N. Wood & Y. A. Breyer (Eds.), *Success in higher education: Transitions to, within and from university* (pp. 269-286). Springer Nature.
https://doi.org/https://doi.org/10.1007/978-981-10-2791-8_16
- Gribble, C., & Blackmore, J. (2012). Re-positioning Australia's international education in global knowledge economies: Implications of shifts in skilled migration policies for universities. *Journal of Higher Education Policy and Management*, 34(4), 341-354.
<https://doi.org/10.1080/1360080X.2012.689181>
- Gue, L. R., & Holdaway, E. A. (1973). English proficiency tests as predictors of success in graduate studies in education. *Language Learning*, 23(1), 89-103.
- Gunn-Lewis, J. (2000). *English entry levels and academic success*. The Fourth Pacific Rim - First Year in Higher Education Conference: Creating Futures for a New Millennium, Queensland University of Technology, Brisbane.
- Haertal, E. H. (1999). Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice*, 18(4), 5-9.
<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-3992.1999.tb00276.x>
- Hagedorn, L. S. (2012). How to define retention: A new look at an old problem. In A. Seidman (Ed.), *College student retention: Formula for student success* (2nd ed., pp. 81-100). Rowman & Littlefield Publishers.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items* (2nd ed.). Lawrence Erlbaum Associates.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. Edward Arnold.
- Hannon, O., Smith, L. R., & Lã, G. (2017). Success at university: The student perspective. In L. N. Wood & Y. A. Breyer (Eds.), *Success in higher education* (pp. 257-268).

Springer.

- Hamp-Lyons, L. (2007). The impact of testing practices on teaching: Ideologies and alternatives. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 487-504). Springer.
https://doi.org/https://doi.org/10.1007/978-0-387-46301-8_35
- Harsch, C., & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing*, 17(4), 228-250. <https://doi.org/https://doi.org/10.1016/j.asw.2012.06.003>
- Harvey, A., & Luckman, M. (2014). Beyond demographics: Predicting student attrition within the Bachelor of Arts degree. *The International Journal of First Year in Higher Education*, 5(1), 19-29. <https://doi.org/https://doi.org/10.5204/intjfyhe.v5i1.187>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Heeren, J., Speelman, D., & Wachter, L. D. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: Implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1458-1473.
<https://doi.org/https://doi.org/10.1080/13670050.2019.1709411>
- Higher Education Standards Framework (Threshold Standards) 2021* (Cth).
<https://www.legislation.gov.au/Details/F2022C00105>
- Higher Education Support Amendment (Job-Ready Graduates and Supporting Regional and Remote Students) Bill 2020* (Cth).
https://parlinfo.aph.gov.au/parlInfo/download/legislation/billsdgs/7527030/upload_binary/7527030.pdf;fileType=application%2Fpdf
- Hill, B., Winmar, G., & Woods, J. (2020). Exploring transformative learning at the cultural interface: Insights from successful Aboriginal university students. *The Australian Journal of Indigenous Education*, 49(1), 2-13.
<https://doi.org/https://doi.org/10.1017/jie.2018.11>
- Hill, K., Storch, N., & Lynch, B. (1999). *A comparison of IELTS and TOEFL as predictors of academic success*. International English Language Testing System (IELTS) Research Reports. https://www.ielts.org/-/media/research-reports/ielts_rr_volume02_report3.ashx
- Ho, D. Y. F., & Spinks, J. A. (1985). Multivariate prediction of academic performance by Hong Kong University students. *Contemporary Educational Psychology*, 10(3), 249-

259. [https://doi.org/10.1016/0361-476X\(85\)90021-9](https://doi.org/10.1016/0361-476X(85)90021-9)
- Holder, G. M., Jones, J., Robinson, R. A., & Krass, I. (1999). Academic literacy skills and progression rates amongst pharmacy students. *Higher Education Research & Development, 18*(1), 19-30. <https://doi.org/10.1080/0729436990180103>
- Holliday, A. (2010). Analysing qualitative data. In B. Paltridge & A. Phakiti (Eds.), *Continuum Companion to Research Methods in applied Linguistics* (pp. 98-110). Continuum International.
- Hoyne, G. F., & McNaught, K. (2016). Changing practices to better support first-year Health Science students. Students Transitions Achievement Retention & Success (STARS). <https://unistars.org/papers/STARS2016/01A.pdf>
- Hughes, A. (1991). *Testing for language teachers*. Cambridge University Press.
- Huijser, H., Kimmins, L., & Galligan, L. (2008). Evaluating individual teaching on the road to embedding academic skills. *Journal of Academic Language & Learning, 2*(1), A23-A38. <http://www.journal.aall.org.au/index.php/jall/article/view/61/54>
- Humphreys, P. (2015). *English language proficiency in higher education: Student conceptualisations and outcomes* [Doctoral thesis, Griffith University]. https://research-repository.griffith.edu.au/bitstream/handle/10072/367494/Humphreys_2016_01Thesis.pdf?sequence=1
- Humphreys, P. (2017). EMI in anglophone nations: Contradiction in terms or cause for consideration? In B. Fenton-Smith, P. Humphreys, & I. Walkinshaw (Eds.), *English Medium Instruction in Higher Education in Asia Pacific: From policy to pedagogy* (pp. 93-114). Springer.
- Humphreys, P. (2022). The need for English language development in anglophone higher education settings: A case study from Australia. *Journal of English for Academic Purposes, 59*, 1-8. <https://doi.org/https://doi.org/10.1016/j.jeap.2022.101156>
- Humphreys, P., & Gribble, C. (2013). *Outcomes: English language and the transition to work or further study* Five years on: English language competence of international students (pp. 75-104). International Education Association of Australia. <https://www.ieaa.org.au/documents/item/54>
- Humphreys, P., Haugh, M., Fenton-Smith, B., Lobo, A., Michael, R., & Walkinshaw, I. (2012). *Tracking international students' English proficiency over the first semester of undergraduate study*. IELTS Research Reports Online Series. <https://www.ielts.org/teaching-and-research/research-reports/online-series-2012-1>

- Ihlenfeldt, S. D., & Rios, J. A. (2022). A meta-analysis on the predictive validity of English language proficiency assessments for college admissions. *Language Testing*, 40(2), 276-299. <https://doi.org/10.1177/02655322221112364>
- Im, G.-H., Shin, D., & Cheng, L. (2019). Critical review of validation models and practices in language testing: their limitations and future directions for validation research. *Language Testing in Asia*, 9(1), 1–26. <https://doi.org/10.1186/s40468-019-0089-4>
- Ingram, D., & Bayliss, A. (2007). *IELTS as a predictor of academic language performance, Part 1*. IELTS Research Reports. <https://www.ielts.org/teaching-and-research/research-reports/volume-07-report-3>
- International Language Testing Association. (2020). *International Language Testing Association Guidelines for Practice*. <https://www.iltaonline.com/page/ILTAGuidelinesforPractice>
- Jackson, C. (September 13, 2016). The other side of attrition. *The Australian*. <http://www.theaustralian.com.au/higher-education/the-other-side-of-attrition/news-story/b566045b8f93691a8283ecd8a9e83309>
- James, R., Baldwin, G., Coates, H., Krause, K.-L., & McInnis, C. (2004). *Analysis of equity groups in higher education 1991–2002*. Canberra: Commonwealth of Australia. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.154.8116&rep=rep1&type=pdf>
- Johnes, J. (1997). Inter-university variations in undergraduate noncompletion rates: A statistical analysis by subject of study. *Journal of Applied Statistics*, 24(3), 343-362. <https://doi.org/10.1080/02664769723738>
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133. <https://doi.org/10.1177/1558689806298224>
- Jones, J., Holder, G. M., & Robinson, R. A. (2000). School subjects and academic literacy skills at university. *Australian Journal of Career Development*, 9(2), 27-31.
- Jones, J., Krass, I., Holder, G. M., & Robinson, R. A. (2000). Selecting pharmacy students with appropriate communication skills. *American Journal of Pharmaceutical Education*, 64(1), 68-73. <https://www.proquest.com/docview/211208875/citation/43AE432B4D804236PQ/1?accountid=26503>
- Kane, M. (1992). An argument-based approach to validity. *Quantitative Methods in Psychology*, 112, 527-535.

- Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement*, 2, 135-170.
https://www.tandfonline.com/doi/epdf/10.1207/s15366359mea0203_1?needAccess=true&role=button
- Kane, M. (2011). The errors of our ways. *Journal of Educational Measurement*, 48(1), 12-30.
<https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1745-3984.2010.00128.x>
- Kane, M. (2012). Articulating a validity argument. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 34-47). Routledge.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement* 38(4), 319-342.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Greenwood Publishing.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
<https://onlinelibrary.wiley.com/doi/epdf/10.1111/jedem.12000>
- Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, 18(2), 5-17.
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17-24.
<https://doi.org/10.1037/h0057123>
- Kerstjens, M., & Nery, C. (2000). *Predictive validity in the IELTS test: A study of the reliability between IELTS scores and students' subsequent academic performance*. International English Language Testing System (IELTS) Research Reports.
https://www.ielts.org/-/media/research-reports/ielts_rr_volume03_report4.ashx
- Kift, S. (2009). *Articulating a transition pedagogy to scaffold and to enhance the first year student learning experience in Australian higher education: Final Report for ALTC Senior Fellowship Program*.
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5f64de7c01215c0fea461e539cc4991d9b9979ee>
- Kift, S. M., Butler, D. A., Field, R. M., McNamara, J., Brown, C., & Gamble, N. (2010). *Conceptualising a capstone experience for law students*. Australasian Law Teachers Association 65th Annual Conference, 4-7th July 2010, University of Auckland, Auckland. <https://eprints.qut.edu.au/238695/1/c40301.pdf>
- Kift, S., & Nelson, K. (2005). *Beyond curriculum reform: Embedding the transition*

- experience*. Higher Education Research and Development Society of Australasia annual conference, Melbourne, Australia.
- Kim, A. I. (2021). *Rethinking post-entry language assessment policies in the context of U.S. higher education: A socially responsible approach*. [Doctoral thesis, Georgetown University]. <https://www.proquest.com/docview/2618513198?pq-origsite=gscholar&fromopenview=true>
- Kinash, S., & Crane, L. (2015). *Enhancing graduate employability of the 21st century learner*. International Mobile Learning Festival, Hong Kong. http://epublications.bond.edu.au/fsd_papers/99
- Klebanov, B. B., Ramineni, C., Kaufer, D., Yeoh, P., & Ishizaki, S. (2017). Advancing the validity argument for standardized writing tests using quantitative rhetorical analysis. *Language Testing*, 36(1), 125-144. <https://doi.org/10.1177/0265532217740752>
- Knoch, U. (2007). *Diagnostic writing assessment: The development and validation of a rating scale*. [Doctoral thesis, The University of Auckland]. <https://researchspace.auckland.ac.nz/handle/2292/2343>
- Knoch, U. (2012). At the intersection of language assessment and academic advising: Communicating results of a large-scale diagnostic academic English writing assessment to students and other stakeholders. *Papers in Language Testing and Assessment*, 1(1), 31-49. https://www.altanz.org/uploads/5/9/0/8/5908292/2._plta_1_1__knoch.pdf
- Knoch, U., & Chapelle, C. A. (2017). Validation of rating processes within an argument-based framework. *Language Testing*, 35(4), 477-499. <https://doi.org/https://doi.org/10.1177/0265532217710049>
- Knoch, U., & Elder, C. (2010). Validity and fairness implications of varying time conditions on a diagnostic test of academic English writing proficiency. *System*, 38, 63-74. <https://doi.org/doi:10.1016/j.system.2009.12.006>
- Knoch, U., & Elder, C. (2013). A framework for validating post-entry language assessments (PELAs). *Papers in Language Testing and Assessment*, 2(2), 48-66. <https://doi.org/10.58379/YZLQ8816>
- Knoch, U., & Elder, C. (2016). Post-entry english language assessments at university: How diagnostic are they? In V. Aryadoust & J. Fox (Ed.), *Trends in Language Assessment Research and Practice* (pp. 210-230). Cambridge Scholars Publishing.
- Knoch, U., Elder, C., & O'Hagan, S. (2016). Examining the validity of a post-entry screening tool embedded in a specific policy context. In J. Read (Ed.), *Post-admission language*

- assessment of university students* (pp. 23-42). Springer International.
- Knoch, U., Elder, C., Woodward-Kron, R., Manias, E., Flynn, E., McNamara, T., Huisman, A., & Zhang, B. Y. (2020). Capturing domain expert perspectives in devising a rating scale for a health specific writing test: How close can we get? *Assessing Writing*, *46*, 1-13. <https://doi.org/https://doi.org/10.1016/j.asw.2020.100489>
- Knoch, U., Fairbairn, J., Myford, C., & Huisman, A. (2018). Evaluating the relative effectiveness of online and face-to-face training for new writing raters. *Papers in Language Testing and Assessment*, *7*(1), 61-86.
https://www.altanz.org/uploads/5/9/0/8/5908292/7_1_knoch_et_al.pdf
- Knoch, U. & Frost, K. (2017). *Setting Empirical Standards on the Diagnostic English Language Needs Assessment (DELA)*.
https://arts.unimelb.edu.au/__data/assets/pdf_file/0007/2380885/DELA-standard-setting-report-FINAL.pdf
- Knoch, U., & Macqueen, S. (2019). *Assessing English for Professional Purposes*. Routledge.
- Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, *12*(1), 26-43.
<https://doi.org/https://doi.org/10.1016/j.asw.2007.04.001>
- Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing*, *28*, 39-52.
<https://doi.org/http://dx.doi.org/10.1016/j.jslw.2015.02.005>
- Koramannil, G. (2016). Looking for the invisible: The case of EALD Indigenous students in higher education. *Journal of Academic Language and Learning*, *10*, A87–A100.
<http://journal.aall.org.au/index.php/jall/article/viewArticle/402>.
- Krause, K.-L. (2005). Serious thoughts about dropping out in first year: Trends, patterns and implications for higher education. *Studies in Learning, Evaluation, Innovation and Development*, *2*(3), 55-68. <https://research-repository.griffith.edu.au/bitstream/handle/10072/15410/45960.pdf?sequence=1>
- Krause, K.-L., Hartley, R., James, R., & McInnis, C. (2005). *The first year experience in Australian universities: Findings from a decade of national studies*. Department of Education, Science and Training, Australian Government.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537–67. <https://doi.org/10.1146/annurev.psych.50.1.537>
- Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The Measurement of Attitudes. In D.

- Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 21–76). Lawrence Erlbaum Associates Publishers.
- Kung, T. H., Cheatham, M., ChatGPT, Medenilla, A., Sillos, C., Leon, L. D., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*, 2(2), 1-12. <https://doi.org/https://doi.org/10.1101/2022.12.19.22283643>
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests. A teacher's book*. McGraw-Hill.
- Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance: Methods, theory, and applications*. Academic Press.
- Language Testing Research Centre (n.d.). *Diagnostic English Language Assessment (DELA)*. <https://arts.unimelb.edu.au/language-testing-research-centre/research/projects/dela>
- Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157-172. <https://doi.org/https://doi.org/10.1080/03075079812331380364>
- Lea, M. R., & Street, B. V. (2006). The "Academic Literacies" model: Theory and applications. *Theory Into Practice*, 45(4), 368-377. https://doi.org/https://doi.org/10.1207/s15430421tip4504_11
- Leigh, A. (2021). Richest 10pc three to five times more likely to attend university. *Australian Financial Review*. <https://www.afr.com/work-and-careers/education/richest-10pc-three-to-five-times-more-likely-to-attend-university-20211205-p59ewm>
- Li, J., & Barnard, R. (2011). Academic tutors' beliefs about and practices of giving feedback on students' written assignments: A New Zealand case study. *Assessing Writing*, 16(2), 137-148. <https://doi.org/https://doi.org/10.1016/j.asw.2011.02.004>
- Li, Z. (2015). *An argument-based validation study of the English Placement Test (EPT) – Focusing on the inferences of extrapolation and ramification*. [Doctoral thesis, Iowa State University]. <https://www.proquest.com/docview/1733954827?pq-origsite=gscholar&fromopenview=true>
- Light, R. L., Xu, M., & Mossop, J. (1987). English proficiency and academic performance of international students. *TESOL Quarterly*, 21(2), 251-261. <https://doi.org/https://doi-org.ezproxy.bond.edu.au/10.2307/3586734>
- Lillis, T., & Scott, M. (2007). Defining academic literacies research: Issues of epistemology, ideology and strategy. *Journal of Applied Linguistics*.

- <https://doi.org/http://dx.doi.org/doi:10.1558/japl.v4i1.5>
- Linden, K., Teakel, S., & Van der Ploeg, N. (2022). Improving student success with online embedded tutor support in first-year subjects. *Student Success*, 13(2), 42-50.
<https://doi.org/https://doi.org/10.5204/ssj.2338>
- Lloyd-Jones, G., Neame, C., & Medaney, S. (2011). *A multiple case study of the relationship between the indicators of students' English language competence on entry and students' academic progress at an international postgraduate university*.
<https://www.ielts.org/for-researchers/research-reports/volume-11-report-3>
- Lobo, A. (2012). Will we meet again?: Examining the reasons why students are leaving first year university courses and moving towards an approach to stop them. *The International Journal of Learning*, 18(7), 199 - 212.
- Lockwood, J. (2013). The Diagnostic English Language Tracking Assessment (DELTA) writing project: A case for post-entry assessment policies and practices in Hong Kong universities. *Papers in Language Testing and Assessment*, 2, 30-49.
http://www.altanz.org/uploads/5/9/0/8/5908292/3_lockwood.pdf
- Long, M., Ferrier, F., & Heagney, M. (2006). *Stay, play or give it away? Students continuing, changing or leaving university study in first year*. Monash University, Centre for the Economics of Education and Training. <https://files.eric.ed.gov/fulltext/ED505791.pdf>
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, 19(3), 227-336.
<https://journals.sagepub.com/doi/epdf/10.1191/0265532202lt230oa>
- Lydster, C., & Brown, S. (2017). The value of post-entry language assessment (PELA): Outcomes from a first semester undergraduate subject. *Journal of Academic Language and Learning*, 11(1), A39-A57.
<https://journal.aall.org.au/index.php/jall/article/view/434/268>
- Lydster, C., & Murray, J. (2018). Understanding the challenges, yet focusing on the successes: An investigation into Indigenous university students' academic success. *The Australian Journal of Indigenous Education*, 48(2), 107-118.
<https://doi.org/10.1017/jie.2018.15>
- Lydster, C., & Murray, J. (2019). "Not just a tutor": Successful supplementary tuition for Australian Indigenous students in higher education. *Journal of Academic Language and Learning*, 13(1), A140-A160.
<https://journal.aall.org.au/index.php/jall/article/view/609>
- Malczewska-Webb, B. (2016). International Students in Australia: What Makes Them

- Happy? Student Data from the Positive Education Perspective. In D. Gabryś-Barker & D. Galajda (Eds.), *Positive Psychology Perspectives on Foreign Language Learning and Teaching* (pp.193-210). Springer International Publishing.
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. Routledge.
- Martirosyan, N. M., Hwang, E., & Wanjohi, R. (2015). Impact of English proficiency on academic performance of international students. *Journal of International Students*, 5(1), 60-71. <https://doi.org/https://doi.org/10.32674/jis.v5i1.443>
- Matsumoto, M. (2011). Second language learners' motivation and their perception of their teachers as an affecting factor. *New Zealand Studies in Applied Linguistics*, 17(2), 37-52. <https://search.informit.org/doi/abs/10.3316/informit.808628981621588>
- Mauriyat, A. (2021). Authenticity and validity of the IELTS writing test as predictor of academic performance. *Project (Professional Journal of English Education)*, 4(1), 105-115. <https://doi.org/10.22460/project.v4i1.p105-115>
- McKenzie, K., & Schweitzer, R. D. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. *Higher Education Research & Development*, 20, 21-33. <https://doi.org/https://doi.org/10.1080/07924360120043621>
- McKenzie, K., Gow, K., & Schweitzer, R. (2004). Exploring first-year academic achievement through structural equation modelling. *Higher Education Research & Development*, 23(1), 95-112. <https://doi.org/https://doi.org/10.1080/0729436032000168513>
- McLeod, M. (2012). *Looking for an ounce of prevention: The potential for diagnostic assessment in academic acculturation* [Master's thesis, Carleton University]. <https://doi.org/10.22215/etd/2012-06750>
- McMillan, J. (2005). *Course change and attrition from higher education*. Longitudinal Surveys of Australian Youth, Australian Council for Educational Research. https://research.acer.edu.au/lsey_research/43
- McNamara, T. F. (1996). *Measuring second language performance*. Longman.
- McNamara, T. F., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing.
- McNaught, K. & Shaw G. (2016). Preparing undergraduate students to be successful writers: Exploring the spelling, grammar and punctuation skills of students identified as potentially "at risk." *Journal of Academic Language and Learning*, 10(2). <https://journal.aall.org.au/index.php/jall/article/view/406>

- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*. American Council on Education/Macmillan Publishing.
- Morgan, G. B., Zhu, M., Johnson, R. L., & Hodge, K. J. (2014). Interrater reliability estimators commonly used in scoring language assessments: A Monte Carlo investigation of estimator accuracy. *Language Assessment Quarterly*, *11*(3), 304-324. <https://doi.org/10.1080/15434303.2014.937486>
- Mthimunye, K., & Daniels, F. M. (2019). Predictors of academic performance, success and retention amongst undergraduate nursing students: A systematic review. *South African Journal of Higher Education*, *33*(1), 200-220. <https://doi.org/https://doi.org/10.20853/33-1-2631>
- Mulligan, A. C. (1966). Evaluating foreign credentials. *College and University*, *41*(3), 307-313.
- Mumford, S., & Atay, D. (2021). Teachers' perspectives on the causes of rater discrepancy in an English for Academic Purposes context. *Assessing Writing*, *48*, 1-11. <https://doi.org/https://doi.org/10.1016/j.asw.2021.100527>
- Murray, D., & Arkoudis, S. (2013). *Preparation and selection*. Five years on: English language competence of international students (pp. 23-51). International Education Association of Australia. <https://www.ieaa.org.au/documents/item/54>
- Murray, N. (2010). Conceptualising the English language needs of first year university students. *The International Journal of the First Year in Higher Education*, *1*(1), 55-64. <https://fyhejournal.com/article/download/19/48/19-1-140-2-10-20100802.pdf>
- Murray, N. (2012). Ten 'Good Practice Principles' ... ten key questions: considerations in addressing the English language needs of higher education students. *Higher Education Research & Development*, *31*(2), 233-246. <https://doi.org/10.1080/07294360.2011.555389>
- Murray, N. (2014). Reflections on the implementation of post-enrolment English language assessment. *Language Assessment Quarterly* *11*(3), 325-337. <https://doi.org/10.1080/15434303.2013.824975>
- Murray, N. (2016). *Standards of English in Higher Education*. Cambridge University Press.
- Murray, N. (2018). Understanding student participation in the internationalised university: Some issues, challenges, and strategies. *Education Sciences*, *8*(3), 96-106. <https://doi.org/doi:10.3390/educsci8030096>
- Murray, N. (2022). A model to support the equitable development of academic literacy in institutions of higher education. *Journal of Further and Higher Education*, *46*(8),

- 1054-1065. <https://doi.org/https://doi.org/10.1080/0309877X.2022.2044019>
- Nakata, M., Nakata, V., & Chin, M. (2008). Approaches to the academic preparation and support of Australian Indigenous students for tertiary studies. *The Australian Journal of Indigenous Education*, 37(S1), 137–145.
<https://doi.org/10.1375/S1326011100000478>
- Nakata, M., Nakata, V., Day, A., & Martin, G. (2018). Indigenous undergraduates' use of supplementary tutors: Developing academic capabilities for success in higher education studies. *The Australian Journal of Indigenous Education*, 48(2), 119-128.
<https://doi.org/10.1017/jie.2017.39>
- Nelson, K. J., Quinn, C., Marrington, A., & Clarke, J. A. (2012). Good practice for enhancing the engagement and success of commencing students. *Higher Education*, 63, 83-96.
<https://doi.org/10.1007/s10734-011-9426-y>
- Norton, A. (2022). The inequity of Job-ready Graduates for students must be brought to a quick end. Here's how. *The Conversation*. <https://theconversation.com/the-inequity-of-job-ready-graduates-for-students-must-be-brought-to-a-quick-end-heres-how-183808>
- O'Hagan, S. R., & Wigglesworth, G. (2015). Who's marking my essay? The assessment of non-native-speaker and native-speaker undergraduate essays in an Australian higher education context. *Studies in Higher Education*, 40(9), 1729-1747.
<https://doi.org/https://doi.org/10.1080/03075079.2014.896890>
- O'Loughlin, K., & Arkoudis, S. (2009). *Investigating IELTS exit score gains in higher education*. IELTS Research Reports. https://www.ielts.org/-/media/research-reports/ielts_rr_volume10_report3.ashx
- O'Mahony, B., Verezub, E., Dalrymple, J., & Bertone, S. (2013). An evaluation of research students' writing support intervention. *Journal of International Education in Business*, 6(1), 22-34. doi:10.1108/18363261311314935
- Organisation for Economic Co-operation and Development. (2017). *In-depth analysis of the labour market relevance and outcomes of higher education systems: Analytical framework and country practices report*. Enhancing Higher Education System Performance. <https://www.oecd.org/education/skills-beyond-school/LMRO%20Report.pdf>
- O'Sullivan, B., & Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language testing: Theories and practices* (pp. 13-32). Palgrave Macmillan.
- Ozga, J., & Sukhnandan, L. (1998). Undergraduate non-completion: Developing an

- explanatory model. *Higher Education Quarterly*, 52(3), 316-333.
<https://doi.org/10.1111/1468-2273.00100>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578. <https://www.jstor.org/stable/1170653>
- Palmer, L., Levett-Jones, T., Smith, R., & McMillan, M. (2014). Academic literacy diagnostic assessment in the first semester of first year at university. *The International Journal of First Year in Higher Education*, 5(1), 67-78.
<https://doi.org/10.5204/intjfyhe.v5i1.201>
- Paton, M. J. (2007). Why international students are at greater risk of failure: An inconvenient truth. *The Journal of Diversity in Organisations, Communities & Nations*, 6(6), 101-111.
- Pearson, W. S. (2021). The predictive validity of the academic IELTS test: A methodological synthesis. *International Journal of Applied Linguistics*, 172(1), 85-120.
<https://doi.org/10.1075/itl.19021.pea>
- Perry, F. L. (2011). *Research in applied linguistics: Becoming a discerning consumer* (2nd ed.). Routledge.
- Pilcher, N., & Richards, K. (2017). Challenging the power invested in the International English Language testing System (IELTS): Why determining 'English' preparedness needs to be undertaken within the subject context. *Power and Education*, 9(1), 3-17.
<https://doi.org/10.1177/1757743817691995>
- Plakans, L. (2012). Assessment of integrated skills. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 205–212). Blackwell.
- Power, C. N., Robertson, F., & Baker, M. (1987). *Success in higher education*. Commonwealth Tertiary Education, Commission, Australian Government Publishing Service.
- Puddey, I. B., & Mercer, A. (2014). Predicting academic outcomes in an Australian graduate entry medical programme. *BMC Medical Education*, 14(31), 1-12.
<https://bmcmmededuc.biomedcentral.com/articles/10.1186/1472-6920-14-31>
- Podorova, A. (2016). Academic Language Feedback toolkit: Making progress with post-entry language skills development. *Journal of Academic Language and Learning*, 10(1), A141-A154. <http://journal.aall.org.au/index.php/jall/article/viewArticle/401>
- Ransom, L. (2009). Implementing the post-entry English language assessment policy at the University of Melbourne: Rationale, processes, and outcomes. *Journal of Academic Language and Learning* 3(2), A13-A25.

- <http://journal.aall.org.au/index.php/jall/article/viewArticle/71>
- Rea, L., & Parker, A. (2014). *Designing and conducting survey research: A comprehensive guide* (4th ed.), John Wiley & Sons.
- Read, J. (2008). Identifying academic language needs through diagnostic assessment. *Journal of English for Academic Purposes*, 7(3), 180-190.
<https://doi.org/https://doi.org/10.1016/j.jeap.2008.02.001>
- Read, J. (2010). Researching language testing and assessment. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research in Applied Linguistics* (pp. 286-300). Continuum International Publishing Group.
- Read, J. (2015a). *Assessing English proficiency for university study*. Palgrave Macmillan.
- Read, J. (2015b). Issues in post-entry language assessment in English-medium universities. *Language Teaching*, 48(2), 217-234. doi:10.1017/S0261444813000190
- Read, J. (2016a). *Post-admission language assessment of university students*. Springer.
- Read, J. (2016b). Some key issues in post-admission language assessment. In J. Read (Ed.), *Post-admission language assessment of university students* (pp. 3-22). Springer.
- Read, J. (2019). Postentry English language assessment in universities. In G. X. (Ed.), *Second Handbook of English Language Teaching* (pp. 395-414). Springer.
https://doi.org/https://doi.org/10.1007/978-3-030-02899-2_26
- Read, J., & von Randow, J. (2013). A university post-entry English language assessment: Charting the changes. *International Journal of English Studies*, 13(2), 89-110.
<https://eric.ed.gov/?id=EJ1042859>
- Reynolds, A. L., & Weigand, M. J. (2010). The relationships among academic attitudes, psychological attitudes, and the first-semester academic achievement of first-year college students. *Journal of Student Affairs Research and Practice*, 47(2), 175-195.
<https://doi.org/https://doi.org/10.2202/1949-6605.6004>
- Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics: Quantitative, qualitative. and mixed-methods research*. Routledge.
- Riazi, A. M. & Candlin, C. N. (2014) Mixed-methods research in language teaching and learning: Opportunities, issues and challenges. *Language Teaching*, 47(2), 135-173.
<https://doi.org/10.1017/S0261444813000505>
- Riazi, M. (2013). Concurrent and predictive validity of Pearson Test of English Academic (PTE Academic). *Papers in Language Testing and Assessment*, 2(2), 1-27.
http://www.altanz.org/uploads/5/9/0/8/5908292/2_riazi.pdf
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university

- students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353-387. <https://doi.org/10.1037/a0026838>
- Rocheouste, J., Oliver, R., Mulligan, D., & Davies, M. (2011). The English Language Growth Project. *The Australian Universities' Review*, 53(1), 102–104. <https://search.informit.org/doi/pdf/10.3316/informit.806137849663124>
- Rocheouste, J., Oliver, R., Bennell, D., Anderson, R., Cooper, I., & Forrest, S. (2016). Teaching Australian Aboriginal higher education students: What should universities do? *Studies in Higher Education*, 42(11), 1-19. doi:10.1080/03075079.2015.1134474.
- Roddan, M. (2002). The determinants of student failure and attrition in first year computing science. *Computing Science, Glasgow University*. <https://www.psy.gla.ac.uk/~steve/localed/docs/roddanpsy.pdf>
- Roessingh, H., & Kover, P. (2002). Working with younger-arriving ESL learners in high school English: Never too late to reclaim potential. *Canadian Modern Language Review*, 58(2), 203-222. <https://teslcanadajournal.ca/index.php/tesl/article/view/926/745>
- Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463-481. <https://doi.org/10.1177/0265532210394633>
- Roever, C., & Phakiti, A. (2018). *Quantitative methods for second language research*. Routledge.
- Roever, C., Fraser, C., & Elder, C. (2014). *Testing ESL sociopragmatics* (Vol. 35). Peter Lang Edition.
- Ruegg, R., Petersen, N., Hoang, H., & Marianne. (2020). Effects of pathways into university on the academic success of international undergraduate students. *Higher Education Research & Development*, 40(6), 1283-1297. <https://doi.org/https://doi.org/10.1080/07294360.2020.1804336>
- Rummel, S. (2020). The value of raters' comments on the writing component of a diagnostic assessment for language advising. *VNU Journal of Foreign Studies*, 36(4), 113-130. [https://doi.org/ https://doi.org/10.25073/2525-2445/vnufs.4579](https://doi.org/https://doi.org/10.25073/2525-2445/vnufs.4579)
- Rummel, S., & Bitchener, J. (2015). The effectiveness of written corrective feedback and the impact Lao learners' beliefs have on uptake. *Australian Review of Applied Linguistics*, 38(1), 66-84. <https://doi.org/https://doi.org/10.1075/ara1.38.1.04rum>
- Scott, D., & Smart, W. (2005). *What factors make a difference to getting a degree in New Zealand?* https://www.educationcounts.govt.nz/__data/assets/pdf_file/0007/9547/modelling-

completion-factors-report-sfs.pdf

- Scouller, K., Bonanno, H., Smith, L., & Krass, I. (2008). Student experience and tertiary expectations: Factors predicting academic literacy amongst first-year pharmacy students. *Studies in Higher Education, 33*(2), 167-178.
<https://doi.org/10.1080/03075070801916047>
- Shah, C., & Burke, G. (1996). *Student flows in Australian higher education*. A. G. P. Service.
<https://files.eric.ed.gov/fulltext/ED425671.pdf>
- Shohamy, E., Gordon, C. M., & Kraemer, R. (1992). The effect of raters' background and training on the reliability of direct writing tests. *The Modern Language Journal, 76*(1), 27-33. <https://www.jstor.org/stable/329895>
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453.
<https://doi.org/10.3102/00346543075003417>
- Smith, M., Therry, L., & Whale, J. (2012). Developing a model for identifying students at risk of failure in a first year accounting unit. *Higher Education Studies, 2*(4), 91-102.
<https://doi.org/10.5539/hes.v2n4p91>
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation, 9*(4), 1-11.
<https://doi.org/10.7275/96jp-xz07>
- Stevenson, M. D., & Kokkinn, B. A. (2009). Evaluating one-to-one sessions of academic language and learning. *Journal of Academic Language & Learning, 3*(2), A36-A50.
<http://journal.aall.org.au/index.php/jall/article/viewFile/86/66>
- Storch, N., & Hill, K. (2008). What happens to international students' English after one semester at university? *Australian Review of Applied Linguistics, 31*(1), 4.1-4.17.
<https://doi.org/10.2104/aral0804>
- Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes, 8*(3), 207-223.
<https://doi.org/10.1016/j.jeap.2009.03.001>
- StudyAssist. (n.d.). *Low completion rate*. <https://www.studyassist.gov.au/help-loans-commonwealth-supported-places-csps/low-completion-rate>
- Strickland, V. P. (1998). *Attendance and grade point average: A study*.
<https://files.eric.ed.gov/fulltext/ED423224.pdf>
- Sugimoto, R. A. (1966). *The relationship of selected predictive variables to foreign student*

- achievement at the University of California, Los Angeles* University of Southern California. <https://www.proquest.com/docview/302233683?pq-origsite=gscholar&fromopenview=true>
- Syron, M., & McLaughlin, J.M. (2010). Indigenous knowledges: Informing and supporting Indigenous students during their first year at university. *Proceedings of the 13th Pacific Rim First Year in Higher Education Conference 2010*, QUT Publications, 1-11. <http://eprints.qut.edu.au/42353/>
- Tanaka, K., & Ellis, R. (2003). Study-abroad, language proficiency and learner beliefs about language learning. *Japan Association for Language Teaching*, 25(1), 63-85. <https://doi.org/10.37546/JALTJJ25.1-3>
- Tertiary Collection of Student Information. (n.d.a). *Attrition rate*. <https://www.tcsisupport.gov.au/node/7770>
- Tertiary Collection of student Information. (n.d.b). *Retention rate*. <https://www.tcsisupport.gov.au/glossary/glossaryterm/Retention%20rate>
- Tertiary Collection of Student Information. (n.d.c). *Success rate*. <https://www.tcsisupport.gov.au/node/8013>
- Tertiary Education Quality and Standards Agency. (2012). *English language standards for higher education*. <https://www.aall.org.au/publications/english-standards-for-higher-education-elshe-2012/>
- Tertiary Education Quality and Standards Agency. (2020). *Good practice note: Improving retention and completion of students in Australian higher education*. <https://www.teqsa.gov.au/sites/default/files/good-practice-note-improving-retention-completion-students-v2-0.pdf>
- The University of Melbourne (n.d.). *Understanding your DELA results*. <https://students.unimelb.edu.au/academic-skills/home/english-language-development-DELA/understanding-your-dela-results>
- Thomas, D. T., Walsh, E. T., Torr, B. M., Alvarez, A. S., & Malagon, M. C. (2021). Incorporating high-impact practices for retention: A learning community model for transfer students. *Journal of College Student Retention: Research, Theory & Practice*, 23(2), 243–263. <https://doi.org/10.1177/1521025118813618>
- Times Higher Education. (2023). *Bond University*. <https://www.timeshighereducation.com/world-university-rankings/bond-university>
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). The University of Chicago Press.

- Tinto, V., & Love, A. G. (1995). *A longitudinal study of learning communities at LaGuardia Community College*. <https://files.eric.ed.gov/fulltext/ED380178.pdf>
- Toulmin, S. E. (1958). *The uses of argument*. University Press
- Toulmin, S. E. (2003). *The uses of argument (updated edition)*. Cambridge University Press.
- Trotter, E., & Roberts, C. A. (2007). Enhancing the early student experience. *Higher Education Research & Development*, 25(4), 371–386.
<https://doi.org/https://doi.org/10.1080/07294360600947368>
- Tumen, S., Shulruf, B., & Hattie, J. (2008). Student pathways at the university: Patterns and predictors of completion. *Studies in Higher Education*, 33(3), 233-252.
<https://doi.org/10.1080/03075070802049145>
- Tynan, L., & Johns, K. (2015). Piloting the Post-Entry Language Assessment: Outcomes from a new system for supporting research candidates with English as an additional language. *Quality in Higher Education*, 21(1), 66-78.
<https://doi.org/https://doi.org/10.1080/13538322.2015.1049442>
- Universities Australia. (2017). *Indigenous strategy 2017– 2020*.
<https://www.universitiesaustralia.edu.au/wp-content/uploads/2019/06/Indigenous-Strategy-2019.pdf>
- Universities Australia. (2020). *Indigenous strategy annual report*.
<https://www.universitiesaustralia.edu.au/wp-content/uploads/2020/02/Indigenous-strategy-second-annual-report.pdf>
- Universities Australia. (2022). *2022 Higher education facts and figures*.
https://www.universitiesaustralia.edu.au/wp-content/uploads/2022/09/220207-HE-Facts-and-Figures-2022_2.0.pdf
- Universities Australia. (2023). *International students drive ongoing recovery*.
<https://www.universitiesaustralia.edu.au/media-item/international-students-drive-ongoing-recovery/>
- University of Auckland. (n.d.a). *DELNA*. <https://www.auckland.ac.nz/en/students/student-support/academic-support/delna.html>
- University of Auckland. (n.d.b). *DELNA bibliography*.
<https://www.auckland.ac.nz/en/students/student-support/academic-support/delna/bibliography.html>
- van Dijk, T. (2015). Tried and tested: Academic literacy tests as predictors of academic success. *Tijdschrift voor Taalbeheersing*, 37(2), 159-186. https://www.aup-online.com/docserver/fulltext/15739775/37/2/03_TVT2015.2.VAND.pdf?expires=16

89026078&id=id&accname=guest&checksum=EE0510C0DFFC0BC77CB67E7E41E
A29F1

- van Dyk, T. J. (2005). Towards providing effective academic literacy intervention. *Per Linguam*, 21(2), 38-51. <https://doi.org/https://doi.org/10.5785/21-2-75>
- Veitch, S., & Johnson, S. (2022). Positioning PELA practice within language and literacy development. *Journal of Academic Language & Learning*, 16(1), 140-151. <https://journal.aall.org.au/index.php/jall/article/view/807>
- Vinke, A. A., & Jochems, W. M. G. (1993). English proficiency and academic success in international postgraduate education. *Higher Education*, 26(3), 275-285. <http://www.jstor.org/stable/3447593>
- von Randow, J. (2013). *The DELNA language advisory session: How do students respond? Working together: Planting the seed: The Association of Tertiary Learning Advisors Aotearoa/New Zealand (ATLAANZ)*, Hamilton, New Zealand.
- Voss, E. (2012). *A validity argument for score meaning of a computer-based ESL academic collocational ability test based on a corpus-driven approach to test design* [Doctoral thesis, Iowa State University]. <https://www.proquest.com/docview/1095402129?pq-origsite=gscholar&fromopenview=true>
- Walker-Gibbs, B., Ajjawi, R., Rowe, E., Skourdoumbis, A., Thomas, M. K. E., O'Shea, S., Bennett, S., Fox, B., & Alsen, P. (2019). *Success and failure in higher education on uneven playing fields*. National Centre for Student Equity in Higher Education. <https://www.ncsehe.edu.au/publications/success-failure-higher-education-uneven-playing-fields/>
- Warburton, M. (2021). *The rhetoric and reality of job-ready graduates*. <https://melbourne-cshe.unimelb.edu.au/publications/fellow-voices/the-rhetoric-and-reality-of-job-ready-graduates>
- Watt, D., & Roessingh, H. (2001). The dynamics of ESL dropout: Plus ça change. *Canadian Modern Language Review*, 58(2), 203-222.
- Webb, B. (2011). Australian experience of culturally diverse university classrooms. In J. Arabski & A. Wojtaszek (Eds.), *Aspects of culture in second language acquisition and foreign language learning* (pp.121-137). Springer.
- Webb, B. (2014). Academic discourse and self-efficacy in diverse settings In D. Gabryś-Barker & J. Bielska (Eds.), *Cultural Interfaces in Academic Setting and Beyond* (pp. 35-50). Uniwersytet Śląski.
- Webb, C., English, L., & Bonanno, H. (1995). Collaboration in subject design: Integration of

- the teaching and assessment of literacy skills into a first-year accounting course. *Accounting Education*, 4(4), 335-350.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263-287. <https://doi.org/10.1191/026553298670883954>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Whatman, S., McLaughlin, J., Willsted, S., Tyhuis, A., & Beetson, S. (2008). Quality and efficacy of the Indigenous tutorial assistance scheme (ITAS) for university students. *Australian Journal of Indigenous Education*, 37, 118-130. doi:10.1375/S1326011100000454.
- Whiteford, G., Shah, M., & Sid Nair, C. (2013). Equity and excellence are not mutually exclusive: A discussion of academic standards in an era of widening participation. *Quality Assurance in Education*, 21(3), 299-310. <https://doi.org/10.1108/QAE-Apr-2012-0020>
- Wong, C., Delante, N. L., & Wang, P. (2017). Using PELA to predict international business students' English writing performance with contextualised English writing workshops as intervention program. *Journal of University Teaching & Learning Practice*, 14(1), 1-21. <https://doi.org/10.10/030582/jcu/singapore/2016>
- Woodrow, L. (2006). Academic success of international postgraduate education students and the role of English proficiency. *University of Sydney Papers in TESOL*, 1, 51-70.
- Woodrow, L. (2014). *Writing about quantitative research in applied linguistics*. Palgrave Macmillan.
- Woodward-Kron, R. (2007). Negotiating meanings and scaffolding learning: Writing support for non-English speaking background postgraduate students. *Higher Education Research & Development*, 26(3), 253-268. doi:10.1080/07294360701494286
- Wu-Pong, S., Windridge, G., & Osborne, D. (1997). Evaluation of Pharmacy School Applicants Whose First Language Is Not English. *American Journal of Pharmaceutical Education*, 61(1), 61-66. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=26faaed292525322a3d4c0b81e6d477c5b53448a>
- Yen, D. A., & Kuzma, J. (2009). Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students. *Worcester Journal of Learning and Teaching*, 3, 1-7. <https://eprints.worc.ac.uk/id/eprint/811>
- York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success.

Practical Assessment, Research & Evaluation, 20(5), 1-20. <https://doi.org/https://doi.org/10.7275/hz5x-tx03>

Zaccagnini, M. & Verenikina, I. (2013). Peer Assisted Study Sessions for postgraduate international students in Australia. *Journal of Peer Learning*, 6(2013-14), 86-102. <http://ro.uow.edu.au/ajpl/vol6/iss1/8>

APPENDICES

APPENDIX A: FEATURES OF DIAGNOSTIC ENGLISH LANGUAGE ASSESSMENT (DELA)

Features	Key Questions	DELA (the University of Melbourne)
1. Target	Who is targeted?	Students potentially at risk of academic failure (e.g., students entering with minimum required English test result and students entering through non-traditional pathways, including foundation courses). Available to all students at Melbourne University.
2. Mandatory completion	Is the completion of the PELA compulsory for targeted students?	Mandatory for targeted students. Subsequent language development options are compulsory for students scoring below 3.4 out of 6 overall.
3. Development	Where was the PELA developed and by whom?	Professional language assessors at LTRC, the University of Melbourne University.
4. Target skills	What are the skills targeted by the PELA?	Reading, writing and listening.
5. Mode of delivery	What is the mode of delivery of the PELA?	Pen-and-paper based. Reading (45 minutes): 2 reading passages of approximately 1,500 words each on general interest topics. Question types include cloze, summarizing, matching, transfer of information, multiple choice, true/false, and short answer. Writing (30 minutes): argumentative essay of minimum 250 words. Input text is provided (a set of ideas, in separate sentences) and students are encouraged to use their own ideas.

Features	Key Questions	DELA (the University of Melbourne)
		Listening (30 minutes): A mini lecture on general interest topic divided into four sections. A short reading passage provided before the recording is played.
6. Assessment	How is the PELA assessed and by whom?	<p>Students' performance on each sub-test reported on a 6-point scale.</p> <p>Reading: Students assessed on their skills regarding speed reading, locating specific information, as well as causes and effects, sequences, contrasts, distinguishing between key and supporting points, selecting appropriate words that fit the meaning and grammar text, summarising main topics, concluding based on information in a passage, distinguishing fact from opinion, and reorganising information in a figure of table. A detailed marking key is provided to assess reading and listening.</p> <p>Writing: Trained raters assess students' writing based on three criteria: Grammar and vocabulary, organisation, and content. Points are deducted if students use large sections of the input text. Writing is single marked by human raters on a 6-point Likert type scale.</p> <p>Listening: students are assessed on their ability to locate and recall specific or important information, reorganize information to fit a figure or table, summarise key ideas, distinguish between key and supporting points.</p> <p>For information on DELA results see: https://students.unimelb.edu.au/academic-skills/english-language-development/understanding-your-dela-results</p>

Features	Key Questions	DELA (the University of Melbourne)
7. On-campus support	What support mechanisms is the PELA linked to?	<p>Credit-bearing subjects including Academic English 1 for undergraduates and Presenting Academic Discourse for postgraduates.</p> <p>Academic English tutorials (10 weeks) for postgraduates in science, arts, food science, education, engineering and information technology fields.</p> <p>“Melbourne Talks”, a free 6-week peer assisted support programme. See: https://students.unimelb.edu.au/careers/get-career-ready/leadership-and-employability-programs/student-peer-leader-network/melbourne-talks</p> <p>Academic Skills (ALL unit) workshops. See: https://students.unimelb.edu.au/student-support/advice-and-help/student-services-workshops-and-sessions/upcoming-workshops</p>

APPENDIX B: FEATURES OF MEASURING THE ACADEMIC SKILLS OF UNIVERSITY STUDENTS (MASUS)

Features	Key Questions	MASUS (the University of Sydney)
1. Target	Who is targeted?	Targeted cohorts and/or faculties. Undergraduate and postgraduate.
2. Mandatory completion	Is the completion of the PELA compulsory for targeted students?	Compulsory for cohorts and/or faculties targeted. Subsequent language developments may not be mandatory nor monitored unless recommendations are embedded into subjects.
3. Development	Where was the PELA developed and by whom?	ALL staff at the University of Sydney in 1993. Has been used at other universities including the University of Newcastle, the University of New South Wales, the University of Wollongong.
4. Target skills	What are the skills targeted by the PELA?	Discipline specific essay writing.
5. Mode of delivery	What is the mode of delivery of the PELA?	Discipline specific integrated writing based on data associated with students' study area. Paper based or online. The task can be implemented during subjects' lecture hours to large groups in lecture theatres, in smaller tutorial groups or outside of class time, off campus.
6. Assessment	How is the PELA assessed and by whom?	Essays assessed on a 4-point Likert type scale on various criteria categorised into four themes: (A) use of input material, (B) structure and development of text, (C) control of academic style and (D) grammatical correctness.

Features	Key Questions	MASUS (the University of Sydney)
		Raters vary depending on the cohort or faculty. Rating sheets are provided for both language assessment experts and discipline experts.
7. On-campus support	What support mechanisms is the PELA linked to?	<p>Feedback in the form of the rating sheet.</p> <p>Department administering the test makes decisions concerning follow-up action.</p> <p>Examples include workshops or other support for students deemed at risk.</p>

APPENDIX C: FEATURES OF DIAGNOSTIC ENGLISH LANGUAGE NEEDS ASSESSMENT (DELNA)

Features	Key Questions	DELNA (the University of Auckland)
1. Target	Who is targeted?	All first-year undergraduate students and commencing Higher Degree Research (HDR) candidates. DELNA is also used at Monash University with postgraduate Law students and the University of Western Australia with HDR candidates.
2. Mandatory completion	Is the completion of the PELA compulsory for targeted students?	Compulsory for all targeted students.
3. Development	Where was the PELA developed and by whom?	DELNA subtests are from Melbourne University's DELA under a licensing agreement. Created to respond to concerns in the 1990s when there was an increase in the number of EAL students commencing studies. DELNA became operational at the University in the early 2000s.
4. Target skills	What are the skills targeted by the PELA?	Screening stage: vocabulary and reading. Diagnosis stage: academic reading, listening and writing (if required)
5. Mode of delivery	What is the mode of delivery of the PELA?	Computer-based screening task and a paper-based diagnosis. Material that comprises the assessment is of general academic interest no connection to students' planned area of study. Two stages: 1. Screening stage (30 minutes): computer-based, 27-item vocabulary test and 73-item cloze-elide.

Features	Key Questions	DELNA (the University of Auckland)
		<p>2. Diagnostic stage: consisting of listening (30 minutes), reading (45-55 minutes, depending on version) and writing (30 minutes for undergraduates; 70 minutes for HDR). Writing comprises 200-250 words commentary on a provided table or graph (undergraduates) or a 150 word summary of two contrasting texts (part 1) and an academic essay between 250 and 300 words (part 2) on a topic related to the theme of part 1.</p>
6. Assessment	How is the PELA assessed and by whom?	<p>Screening test automatically scored and administered in on-campus computer labs.</p> <p>Screening stage, the classifications include: Good, Satisfactory and Diagnosis required. Cut score of 60 meaning students scoring below 60 required to complete Diagnosis stage.</p> <p>Diagnosis results reported on a scale ranging from 4 (at severe risk of failure, urgent need of support) to 9 (proficient, very unlikely to require language support). Threshold score is 7, with students scoring 5 to 6 considered most at risk academically due to low ELP.</p>
7. On-campus support	What support mechanisms is the PELA linked to?	<p>Based on Diagnostic stage results, students scoring 6 or below requested to attend face-to-face consultation with DELNA Academic English Advisor.</p> <p>Undergraduate students receive feedback on writing based on the three criteria: fluency, content and grammar and vocabular.</p> <p>HDR candidates receive feedback on writing component based on six criteria: Organisation, academic style, quality of discussion, sentence structure, grammar, and vocabulary.</p> <p>Advisors offer suggestions concerning language enhancement opportunities at the University.</p>

APPENDIX D: FEATURES OF POST-ENTRY ASSESSMENT OF ACADEMIC LANGUAGE (PAAL)

Feature	Key Questions	PAAL (the University of Melbourne)
1. Target	Who is targeted?	All incoming undergraduate Commerce and postgraduate Engineering students at Melbourne University (trial). Recommended to all undergraduate and postgraduate students.
2. Mandatory completion	Is the completion of the PELA compulsory for targeted students?	Not compulsory although students are strongly encouraged to take the test.
3. Development	Where was the PELA developed and by whom?	Professional language assessors at LTRC, the University of Melbourne in 2009 based on research and development at LTRC and the University of Auckland. Also used at the University of South Australia, branded as English Language Self-Assessment Tool (ELSAT), used on a non-compulsory basis.
4. Target skills	What are the skills targeted by the PELA?	Vocabulary (15 minutes) Speed reading (10 minutes) Academic essay writing (30 minutes)
5. Mode of delivery	What is the mode of delivery of the PELA?	Completed within 60 minutes online assessment. Can be accessed from anywhere and is not invigilated. Two screening tasks, a speeded cloze-elide task adapted from DELNA, and a C-Test made up of a series of 100-word texts that have letters at the end of every second word missing

Feature	Key Questions	PAAL (the University of Melbourne)
		<p>A 250–300 word academic writing task (argument essay) with a 30-minute time limit (same writing task as DELA writing sub-test)</p> <p>Students complete all three parts of the test, yet only writing of students considered potentially at risk in the screening component is assessed.</p>
6. Assessment	How is the PELA assessed and by whom?	<p>Cloze-elide and C-Test both objectively scored.</p> <p>Scores categorise students into: 1. Proficient, 2. Borderline, 3. At risk.</p> <p>Only students in Borderline and At risk categories have their writing assessed.</p> <p>Trained raters score the writing test using a three-category, analytic rating scale.</p> <p>Students receive report with brief feedback on performance and a language development recommendation.</p>
7. On-campus support	What support mechanisms is the PELA linked to?	Students' writing is accessible online for English tutors to use for diagnostic purposes.

APPENDIX E: EXTERNAL RATERS

In order to obtain objective insight from individuals not associated with the research site, two language assessment experts participated in the study. The two raters are referred to as External rater 1 and External rater 2 in order to deidentify them. The decision to employ two external raters was made to ensure unbiased review of BELA. Employing raters not associated with Bond University to assess approximately 10.2% of the BELA scripts ($n = 120$) and provide feedback on the measure and associated processes reduced any potential conflict and allowed the researcher to evaluate BELA's inter-rater reliability. The total remuneration for both external examiners was \$1,549.64. The external raters are described in more detail below.

External rater 1 has worked as a Teaching English to Speakers of Other Languages (TESOL) educator for two decades. They have taught primarily on English for Academic Purposes (EAP) and on Bridging English programs at an Australian university language centre. External rater 1 has also taught extensively on Cambridge Exam preparation courses for First Certificate of English, Cambridge English Academic and Cambridge English Proficiency. In terms of language assessment, the rater has been involved in assessing writing and speaking on EAP and university pathway programs, including standardisation and moderation sessions and acting as a senior examiner. For 10 years, External rater 1 was an examiner for both IELTS speaking and writing exams. Additionally, the rater has written assessments, produced reading and listening tests for EAP programs at a university language centre and has been an item-writer for the IELTS listening exam.

External rater 2 is an educational test developer and has worked as an educator, examiner, teacher trainer, materials designer and test writer in multiple countries. External rater 2 has extensive experience in both teaching and creating support materials for high school, undergraduate and postgraduate students from culturally and linguistically diverse backgrounds and was co-author of an online IELTS Academic test preparation course. As a test writer and examiner, they have worked on a variety of in-house assessments, as well as large-scale national and international tests, assessing listening, reading, writing, speaking, and use of English. They recently worked with university students as an ALL adviser, assisting students to improve their academic writing skills.

Rater training and moderation was provided to the external raters. Once the external raters had provided informed consent to participate in the study and employment contracts with the University were created, they were provided with a Microsoft OneDrive link to folders containing two pre-recorded videos; one outlined the background of BELA including what it was, how it was used and who the targets for the assessment were, while the second video detailed the task requirements and the rubric, whilst presenting six sample BELA scripts with an accompanying analysis of the ratings given. The four sample BELA scripts were chosen purposefully based on the 86 BELA essays for which there was complete agreement between the original and the second rater (i.e., the researcher). The raters then rated six BELA scripts, again selected based on the 86 BELA essays for which complete agreement between the original raters and the researcher had been established. Raters were provided with a link to an Excel spreadsheet on Microsoft OneDrive to enter ratings into.

After completing rater training online, the researcher interviewed each external rater separately via Microsoft Teams to provide moderation feedback based on the six ratings and gain insight into the raters' initial thoughts about BELA. External rater 1 was provided with a link to a folder on Microsoft OneDrive to 52 version 1 BELA scripts; external rater 2 was provided a link to a separate folder containing 68 version 2 BELA scripts. Within each folder was an Excel spreadsheet in which ratings could be entered. Upon completion of the ratings (4 weeks for external rater 1; 5 weeks for external rater 2), a second semi-structured interview was conducted via Microsoft Teams to clarify any discussion points that had arisen in the initial interview and to gain further, more detailed feedback about BELA and the associated processes from the raters. Each interview lasted approximately 30 minutes each. Each of the four interviews were recorded and transcribed using both Otter (audio) and Microsoft Teams (audio-visual). Any unclear utterances were checked by the researcher by analysing the transcripts in comparison to the recording.

APPENDIX F: INDICATIVE INTERVIEW QUESTIONS FOR EXTERNAL RATERS

Rater training interviews

1. Can you provide any initial feedback on BELA and the associated processes?
2. Please tell me your initial thoughts regarding the scoring rubric.
3. Does the rubric capture relevant aspects of academic essay writing?
4. Is BELA appropriately difficult/challenging?
5. Comparing the two versions of BELA, do you feel they are parallel in design?

Post rating interviews

1. Having assessed approximately 60 BELAs now, can you provide feedback on BELA and the associated processes?
2. Did you feel as though you could implement the scoring procedures consistently? Why/why not?
3. Do you think BELA is sufficient in terms of gauging students' academic essay writing skills?
4. Do you feel the characteristics of the BELA task are similar to what is required of students at university?
5. Does BELA unfairly favour certain groups of test takers?
6. Would there be a stigma attached to poor performance on BELA?
7. Do you think BELA can inform future teaching/learning?

APPENDIX G: BELA VERSION 1

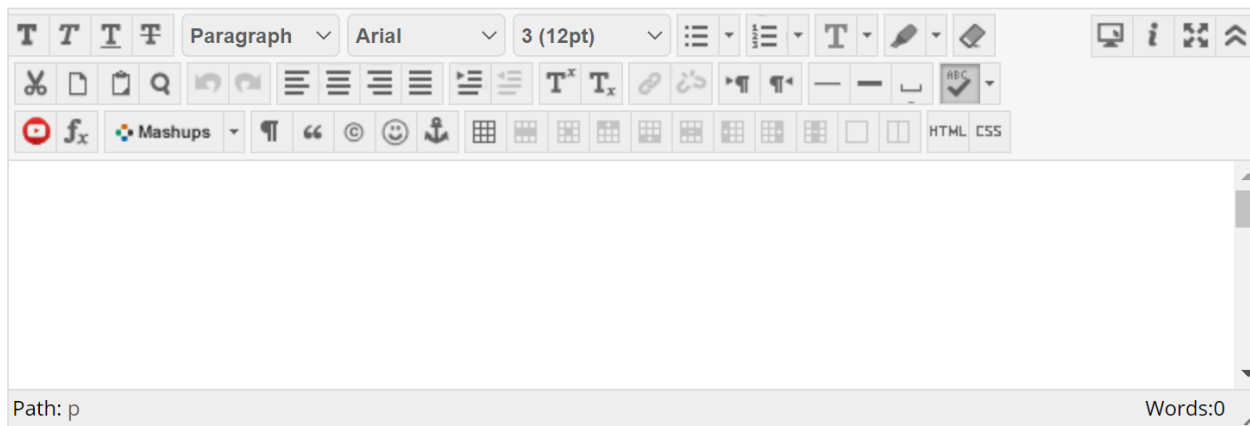
QUESTION 1

9 points [View Rubric](#)

[Save Answer](#)

Write an academic essay on the following subject. You should write between 300 and 400 words and only spend a maximum of 60 minutes on the task.

"Attending classes is far more effective than studying online." Discuss



A rich text editor interface with a toolbar at the top. The toolbar includes icons for text formatting (bold, italic, underline, strikethrough), paragraph alignment (left, center, right, justified), font color, background color, bulleted and numbered lists, indentation, link, unlink, and table insertion. Below the toolbar is a large, empty text area for writing. At the bottom left of the text area, it says "Path: p". At the bottom right, it says "Words:0".

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

[Save All Answers](#)

[Save and Submit](#)

APPENDIX G CONTINUED: BELA VERSION 2

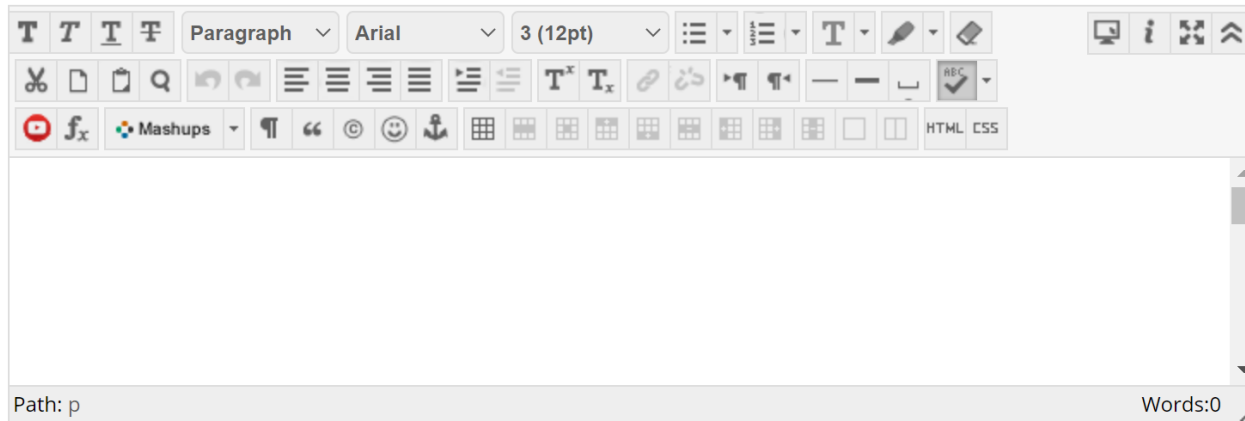
QUESTION 1

9 points [View Rubric](#)

[Save Answer](#)

Write an academic essay on the following subject, You should write between 300 and 400 words and only spend a maximum of 60 minutes on this task

"Exams are the best way to assess a student's understanding of a university subject". Discuss.



The image shows a rich text editor interface. The toolbar includes various icons for text formatting (bold, italic, underline, strikethrough), paragraph alignment (left, center, right, justified), font color, background color, bulleted and numbered lists, indent, decrease/increase indent, link, unlink, table, table border, and other features. The font is set to Arial, size 12pt. The text area is currently empty. At the bottom left of the editor, it says "Path: p" and at the bottom right, it says "Words:0".

Click Save and Submit to save and submit. Click Save All Answers to save all answers.

[Save All Answers](#)

[Save and Submit](#)

APPENDIX H: ANALYTIC RATING SCALE USED WITH BELA

	Below satisfactory	Satisfactory
Organisation	<p>1 (11.111111%)</p> <p>Limited / No introduction or background information given. No clear thesis statement. Limited / no topic sentences. Not logically organised. No conclusion. Poor use of paragraphs.</p>	<p>3 (33.33333%)</p> <p>Introduction provides background. Contains clear thesis statement. Contains satisfactory topic sentences. Logically organised. Contains adequate conclusion. Contains logical paragraphs.</p>
Linking and flow	<p>1 (11.111111%)</p> <p>Inadequate or inaccurate use of cohesive devices. Poor punctuation. Longer sentences tend to be incoherent. Reader has difficulty following the ideas.</p>	<p>3 (33.33333%)</p> <p>Satisfactory use of cohesive devices. Punctuation generally correct, although there may be some omissions or misuse. Longer sentences generally coherent. Reader can follow the ideas without strain.</p>
Grammar & vocabulary	<p>1 (11.111111%)</p> <p>Frequent grammatical errors. Limited range of vocabulary, with frequent errors in word choice or word formation. Mixture of formal and informal styles. Errors cause strain on the reader.</p>	<p>3 (33.33333%)</p> <p>A variety of complex structures. The majority of sentences are error free. Wide range of vocabulary, despite the occasional error in word choice, spelling or word formation. The occasional slip in terms of formal style.</p>

APPENDIX I: RUBRIC FOR MAJOR ESSAY

	Very Good/Excellent	Good	Satisfactory/Aspects missing	Does not meet the minimum requirements
<p>Introduction</p> <p>Background/history</p> <p>Define the problem</p> <p>Thesis Statement</p> <p>Conclusion</p>	<p>Well developed introductory paragraph contains detailed background, a clear explanation or definition of the problem, and a thesis statement clearly stating the position.</p> <p>Conclusion summarises the main topics without repeating previous sentences; writer's opinions and suggestions for change are logical and well thought out.</p>	<p>Introductory paragraph contains some background information and states the problem but does not explain using details. States the thesis of the paper.</p> <p>Conclusion summarises main topics.</p> <p>Some suggestions for change are evident.</p>	<p>Introduction states the thesis but does not adequately explain the background of the problem. The problem is stated, but lacks detail.</p> <p>Conclusion summarises main topics, but is repetitive. No suggestions for change and/or opinions are included.</p>	<p>Thesis and/or problem is vague or unclear.</p> <p>Background details are a seemingly random collection of information, unclear, or not related to the topic.</p> <p>Conclusion does not adequately summarise the main points. No suggestions for change or opinions are included.</p>

<p>Arguments</p> <p>Main Points</p> <p>Body Paragraphs</p> <p>Counter Argument and Rebuttal</p>	<p>Main arguments are well developed with supporting details. Counterargument paragraph(s) acknowledges the opposing view and summarises their main points</p>	<p>Main arguments are present but may lack detail and development in one or two. Counterargument paragraph(s) acknowledges the opposing view but doesn't summarise points.</p>	<p>Main arguments present but all lack development. Counterargument paragraph(s) missing and/or vague</p>	<p>Missing main arguments known for the topic, with poor development of ideas. Counterargument missing, weak or vague</p>
<p>Organisation</p>	<p>Logical, compelling progression of ideas in essay; clear structure which enhances and showcases the central idea or theme and moves the reader through the text. Organisation flows so smoothly the reader hardly thinks about it. Effective, mature, graceful transitions exist throughout the essay.</p>	<p>Overall, the paper is logically developed. Progression of ideas in essay makes sense and moves the reader easily through the text. Strong transitions exist throughout and add to the essay's coherence</p>	<p>Progression of ideas in essay is awkward yet moves the reader through the text without too much confusion. The writer sometimes lunges ahead too quickly or spends too much time on details that do not matter. Transitions appear sporadically, but not equally throughout the essay.</p>	<p>Arrangement of essay is unclear and illogical. The writing lacks a clear sense of direction. Ideas, details or events seem strung together in a loose or random fashion; there is no identifiable internal structure and readers have trouble following the writer's line of thought. Few, forced transitions in the essay or no transitions are present.</p>
<p>Research and referencing</p>	<p>Source material is smoothly integrated into the text. All sources are accurately documented in the</p>	<p>Source material is used. All sources are accurately documented, but a few</p>	<p>Source material is used, but integration may be awkward. All sources are accurately documented, but many are</p>	<p>Lacks sources and/or sources are not accurately documented. Format is incorrect for all sources.</p>

	desired format both in text and in the reference list.	are not in the desired format. Some sources lack credibility.	not in the desired format or lack credibility.	
--	--	---	--	--

APPENDIX J: STUDENT QUESTIONNAIRE



BELA: Student feedback questionnaire

Thank you for agreeing to participate in the current research. The purpose of this study is to gain insight into students' experiences with the BELA process, that is the BELA itself and the connected feedback mechanisms. Please note, by clicking the arrow to proceed, you are consenting to participate.

I am a PhD candidate and this questionnaire will help me to collect data for my doctoral thesis.

I am especially interested in hearing from you, Core 1 students, as the key stakeholder regarding BELA. I do not want to take up too much of your time. Thus, this questionnaire will take approximately 20 minutes to complete.

Your responses are anonymous and will in no way affect your performance in any of your subjects, including Core 1 or your GPA. Your details and responses will be treated confidentially, and they are not traceable to you.

There are no right or wrong answers. I would like you to answer truthfully and to the best of your knowledge. Please respond to each question. That way I will be able to analyse all of the data obtained.

If you experience any distress from participation in this research, please contact: Bond Counselling Services on +61 7 5595 4002.

Should you have any complaints concerning the manner in which this research is conducted, please contact:

Bond University Human Ethics Research Committee
c/o Bond University Office of Research Services.
Bond University, Gold Coast, 4226.
Tel: +61 7 5595 4194 Fax: +61 7 5595 1121 email: buhrec@bond.edu.au

Thank you for taking the time to assist me with this research.

Yours sincerely,

Mr Cameron Lydster



Q1 Did you complete your first Core 1 homework task, BELA (the online essay writing task)?

- Yes (1)
- No (2)

Display This Question:

If Did you complete your first Core 1 homework task, BELA (the online essay writing task)? = No

Q14 Please indicate why you did not complete the first Core 1 homework task, BELA (the online writing task).

Skip To: Q9 If Condition: Please indicate why you did... Is Not Empty. Skip To: Please select your gender:.

Display This Question:

If Did you complete your first Core 1 homework task, BELA (the online essay writing task)? = Yes

Q2 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
The instructions for BELA were clear. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I understood the BELA question. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My writing in BELA was a good example of my academic writing ability. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I had no technical difficulties when completing BELA. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA is a fair task. (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I got help from other sources or copied others' work when completing BELA. (6)

The time limit (60 minutes) was enough time to complete BELA. (7)

I understand the purpose of BELA. (8)

Q3 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
I brainstormed ideas before starting my BELA essay. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I planned my writing prior to starting the writing. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I completed BELA in a quiet environment. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Academic writing is the most important skill at university. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Page Break

Q4 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
BELA is similar to other university assessment tasks. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The academic language I used in BELA was similar to academic language I use in other university tasks. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Some students have an unfair advantage over other students when completing the BELA task. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other skills, such as academic presentations, are equally as important at university as academic writing.

(4)

Page Break

Q5 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
I received the result on my BELA quickly. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My writing was correctly rated. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was easy to access feedback on my BELA. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The feedback I received after completing my BELA has helped me to develop my academic writing. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I received a recommendation/recommendations based on my BELA. (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Academic writing support is available at the University. (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can utilise free support services to develop my academic writing. (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A score of 3 or 5 on BELA indicates that a student's writing is below satisfactory. (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If I didn't receive an email from my lecturer or tutor, it meant my writing was satisfactory. (9)

I do not know if my writing was satisfactory or unsatisfactory. (10)

I felt disadvantaged when completing BELA. (11)

Students who do well on BELA will probably do well on other assessments at university. (12)

A score of 7 or 9 on BELA indicates that a student's academic writing is satisfactory. (13)

Students who do not do well on BELA will probably have a hard time at university. (14)

There is a stigma attached to performing poorly on BELA. (15)

Q6 For the following question, please answer "No" or "Yes".

	No (1)	Yes (2)
Did you attend Student Learning Support to receive feedback on your writing? (1)	<input type="radio"/>	<input type="radio"/>

Display This Question:

If For the following question, please answer "No" or "Yes". = Yes

Q7 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
As a consequence of attending Student Learning Support, my academic writing has improved. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Even though I received feedback and support, my academic writing has not improved. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My English has improved as a result of attending Student Learning Support. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q8 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
Completing BELA has been beneficial for me. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The purpose of BELA is to help students with their academic writing. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA benefits other students. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The goal of BELA is to test students' English. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA was a waste of my time. (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The aim of BELA is to assess students' communication skills. (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The usefulness of
BELA is clear. (7)

○

○

○

○

Page Break

Q9 Please select your gender:

- Male (1)
- Female (2)
- Other (3)

Q10 Please select your current **citizenship(s)**:

Australian (1)

American (US) (2)

Belgian (3)

Brazilian (4)

British (5)

Canadian (6)

Chinese (7)

Colombian (8)

Danish (9)

Dutch (10)

French (11)

Filipino (12)

German (13)

- Indian (14)
- Japanese (15)
- Kiribati (16)
- Malaysian (17)
- New Zealand (18)
- Norwegian (19)
- Omani (20)
- Portuguese (21)
- Russian (22)
- Saudi Arabian (23)
- South Korean (24)
- Swedish (25)
- Swiss (26)
- Taiwanese (27)
- Thai (28)
- Other: (29) _____

Q11 Please select the main **language(s)** you speak at home:

- English (1)
- Mandarin (2)
- Cantonese (3)
- Other Chinese language (e.g., Chang Chow, Hunan, Kan) (4)
- German (5)
- Other German and related languages (6)
- Spanish (7)
- Korean (8)
- French (9)
- Arabic (including Lebanese) (10)
- Hindi (11)
- Bengali (12)
- Telugu (13)
- Other Indian language (14)
- Japanese (15)
- Persian (16)
- Dutch (Netherlandic) (17)
- Tok Pisin (18)
- Malay (19)

- Danish (20)
- Swedish (22)
- Portuguese (23)
- Thai (24)
- Urdu (25)
- Other: (21) _____

Page Break

Q12 Please indicate which **age group** you fall into:

- 16 - 18 (1)
- 19 - 21 (2)
- 22 - 24 (3)
- 25 - 27 (4)
- 28 - 31 (5)
- 32 - 39 (6)
- 40 - 49 (7)
- 50 and above (8)

Page Break

Q13 If you have any additional comments regarding the **BELA process**, please write them here:

End of Block: Default Question Block

APPENDIX K: ACADEMIC STAFF QUESTIONNAIRE



BELA: Core 1 Academic and EAP Professional Staff questionnaire

Thank you for agreeing to participate in the current research. The purpose of this study is to gain insight into experiences with the Bond English Language Assessment (BELA) process, that is the BELA itself and the connected feedback mechanisms.

I am a PhD candidate and this questionnaire will help me to collect data for my doctoral thesis. I believe this study is important, as it will help to present a validity argument for the BELA. I am especially interested in gaining insight from Core 1 and EAP staff.

I do not want to take up too much of your time. Thus, this questionnaire will take approximately 20 minutes to complete.

Your details are anonymous, and your responses will be kept confidential. Survey responses will not be traceable to individuals responding to this questionnaire.

There are no right or wrong answers. I would like you to answer truthfully and to the best of your knowledge. Please respond to each question. That way I will be able to analyse all of the data obtained.

If you experience any distress from participation in this research, please contact: Bond Counselling Services on +61 7 5595 4002.

Should you have any complaints concerning the manner in which this research is conducted, please contact:

Bond University Human Ethics Research Committee
c/o Bond University Office of Research Services.
Bond University, Gold Coast, 4226.
Tel: +61 7 5595 4194 Fax: +61 7 5595 1121 email: buhrec@bond.edu.au

Thank you for taking the time to assist me with this research.

Yours sincerely,

Mr Cameron Lydster



Q1 Please select your role at Bond University:

- Core 1 Academic Staff member (1)
- EAP Teacher at Bond University College (2)

Q2 Approximately how many years have you been in this role?

Page Break

Q3 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
Students plan their writing prior to starting the writing. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students complete BELA in a quiet environment. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students brainstorm ideas before starting their BELA essays. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
A student's writing in BELA is a good example of his/her academic writing ability. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students who do well on BELA will probably do well on other assessments at university. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA scores provide information about a student's academic language ability. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA is similar to other university assessment tasks. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

There would be a correlation between a student's BELA score and his/her score on a high stakes, commercial English language test, such as IELTS. (5)

Q5 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
The BELA task is appropriate for students. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The instructions for BELA are clear. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The criteria (1. organisation, 2. linking and flow, and 3. grammar and vocabulary) capture aspects of performance most relevant to academic essay writing. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Allowing students to complete BELA in their own time is a necessity. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA is a fair task. (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I can consistently
apply the BELA
criteria. (6)

○

○

○

○

It is acceptable
for students to
complete BELA
without the
presence of an
invigilator. (7)

○

○

○

○

A score of 3 or 5 on BELA indicates that a student's academic writing is below satisfactory. (8)

BELA should be completed in a computer lab in the presence of an invigilator. (9)

The time limit (60 minutes) is enough time to complete BELA. (10)

Students plagiarise their BELA. (11)

Students' BELAs are good examples of their academic writing ability. (12)

The nationality or language background of a student influences my rating of his/her BELA. (13)

A score of 7 or 9 on BELA indicates that a student's academic writing is satisfactory. (14)

Because BELA is a low stakes assessment, the chances of students cheating is low. (15)

The content of students' writing in BELA influences my ratings. (16)

Q6 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
Students are given sufficient feedback on their academic writing as part of the BELA process. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is easy for students to access feedback on their BELAs. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There is a stigma attached to performing poorly on BELA. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The feedback students receive after completing BELA helps them to develop their writing skills. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q7 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
I understand the purpose of BELA. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Some students feel disadvantaged when completing BELA. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The purpose of BELA is to help students with their academic writing. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students' writing in BELA can inform instruction. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

As a consequence of attending Student Learning Support, students' academic writing improves. (5)

○ ○ ○ ○

The goal of BELA is to test students' English. (6)

○ ○ ○ ○

It is impossible to ensure all students enrolled in Core 1 complete BELA. (7)

○ ○ ○ ○

BELA is a waste
of students' time.
(8)

Even though
students receive
feedback and
support, their
academic writing
does not improve.
(9)

The usefulness of
BELA is clear.
(10)

The aim of BELA
is to assess
students'
communication
skills. (11)

Students' English
improves as a
result of attending
Student Learning
Support. (12)

Q8 If you have any additional comments regarding the **BELA including the BELA process**, please write them here:

End of Block: Default Question Block

APPENDIX L: BELA USERS QUESTIONNAIRE



BELA users questionnaire

Thank you for agreeing to participate in the current research. The purpose of this study is to gain insight into users' experiences with the Bond English Language Assessment (BELA) process, that is the BELA itself and the connected feedback mechanisms.

I am a PhD candidate and this questionnaire will help me to collect data for my doctoral thesis. I believe this study is important, as it will help to present a validity argument for the BELA. I am especially interested in gaining insight from staff who are provided with students' BELA data.

I do not want to take up too much of your time. Thus, this questionnaire will take approximately 15 minutes to complete.

Your details are anonymous, and your responses will be kept confidential. Survey responses will not be traceable to individuals responding to this questionnaire.

There are no right or wrong answers. I would like you to answer truthfully and to the best of your knowledge. Please respond to each question. That way I will be able to analyse all of the data obtained.

If you experience any distress from participation in this research, please contact: Bond Counselling Services on +61 7 5595 4002.

Should you have any complaints concerning the manner in which this research is conducted, please contact:

Bond University Human Ethics Research Committee
c/o Bond University Office of Research Services.
Bond University, Gold Coast, 4226.
Tel: +61 7 5595 4194 Fax: +61 7 5595 1121 email: buhrec@bond.edu.au

Thank you for taking the time to assist me with this research.

Yours sincerely,

Mr Cameron Lydster



Q1 Please select your role at Bond University:

(if you have multiple roles, please select the one you consider to be your main role)

- Member of the University's Executive (1)
 - Dean/Associate Dean (2)
 - SASQ Manager (3)
 - Director/Manager of a Bond University/Bond University College unit (4)
 - Staff member within the DVC Students and Support Services portfolio (5)
 - Core Curriculum employee (6)
 - BASP Coach/BBS Mentor (7)
 - Other: (8) _____
-

Q2 Approximately how many years have you been in this role?

Q3 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
I understand the purpose of BELA. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am aware of what the BELA task involves. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA results are readily available to me. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I utilise BELA data in my role at the University. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA results are communicated to me in a timely manner. (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BELA is a fair task. (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

A score of 3 or 5 on BELA indicates that a student's academic writing is below satisfactory. (7)

Recommendations provided to students according to their BELA scores are linked to the University's on-campus support. (8)

I understand the meaning of BELA results. (9)

Students whose writing, as measured by BELA, is assessed as being below satisfactory tend to have low academic language proficiency. (10)



Students are given sufficient feedback on their academic writing as part of the BELA process. (11)

○ ○ ○ ○

Students who do not complete BELA tend to perform poorly in the Core 1 subject or discontinue studying. (12)

○ ○ ○ ○

The BELA task is appropriate for students. (13)

○ ○ ○ ○

A score of 7 or 9 indicates that a student's academic writing is satisfactory. (14)

○ ○ ○ ○

The time limit (60 minutes) given is enough time to complete BELA. (15)

I do not need to understand the results of BELA. (16)

The feedback students receive after completing BELA helps them develop their writing. (17)

Q4 For each of the statements below, please select the response which best represents how you feel.

The options are **Strongly Disagree**, **Disagree**, **Agree**, and **Strongly Agree**.

	Strongly Disagree (1)	Disagree (2)	Agree (3)	Strongly Agree (4)
BELA is a useful measure of students' academic writing ability. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Some students feel disadvantaged when completing BELA. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The purpose of BELA is to help students with their academic writing. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The criteria (1. organisation, 2. linking and flow, and 3. grammar and vocabulary) capture aspects of performance most relevant to academic essay writing. (4)

There is a stigma attached to students whose writing is assessed as being below satisfactory. (5)

The goal of BELA is to test students' English. (6)

I do not see the value of BELA. (7)

Feedback from BELA directly informs students' future learning. (8)

As a consequence of attending Student Learning Support, students' academic writing improves. (9)

It is impossible to ensure all students enrolled in Core 1 complete BELA. (10)

The aim of BELA is to assess students' communication skills. (11)

The usefulness of BELA is clear. (12)

All university stakeholders who can influence students' retention are provided with BELA data. (13)

BELA is a waste of students' time. (14)

Completing BELA is beneficial for students. (15)

Q5 If you have any additional comments regarding the **BELA including the BELA process**, please write them here:

End of Block: Default Question Block

APPENDIX M: SUMMARY OF DATA COLLECTION AND ANALYSIS METHODS

	Research question	Data type	Data collection method	Data analysis method
Evaluation inference	RQ1a	1. External rater feedback	Semi-structure interviews	NVivo thematic analysis
	Are BELA scores adequate reflections of observed behaviours?	2. Academic staff questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		3. Inter-rater reliability	180 BELA scripts rated by second rater, 120 by external raters	SPSS analysis (Cohen's kappa coefficient)
		4. Student questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		5. BELA users questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		6. Difficulty and discrimination of test data	Learning management system data	Test difficulty and discrimination analysis

	Research question	Data type	Data collection method	Data analysis method
Generalisability inference	RQ1b	1. External rater feedback	Semi-structure interviews	NVivo thematic analysis
	Does BELA yield results consistent across assessment contexts?	2. BELA version data	Learning management system data	Pearson's chi square test of contingencies
		3. Trial population data	Learning management system data	SPSS analysis (trial population vs study population data)
		4. Academic staff questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		5. Student questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)

	Research question	Data type	Data collection method	Data analysis method
Explanation and Extrapolation inference	RQ1c	1. Academic performance data (BELA, Major Essay, overall grade, GPA)	Learning management system, internal reports	SPSS analysis (Point Biserial correlation)
	Does BELA provide information on test takers' skills, knowledge and characteristics that keep within the understanding of academic ELP?	2. Academic staff questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		3. Common assessment types	Internal analyses	Comparative analysis
		4. External rater feedback	Semi-structure interviews	NVivo thematic analysis
		5. Student questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
	Is BELA an adequate proxy for tasks performed in the academic domain?	6. BELA scores	Learning management system	Statistical analysis of scores based on language background

	Research question	Data type	Data collection method	Data analysis method
Decisions inference	RQ1d	1. Major Essay and GPA data	Learning management system, internal reports	Grade distribution
	Are decisions based on BELA scores appropriate and well communicated?	2. Student questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		3. Academic staff questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		4. BELA users questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		5. ASC feedback survey	Survey Monkey via email	NVivo thematic analysis

	Research question	Data type	Data collection method	Data analysis method
Consequences inference	RQ1e	1. BELA completion rates	Learning management system	Excel analysis
	Are consequences of using BELA and the decisions informed by BELA beneficial to all stakeholders?	2. Student questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		3. Academic staff questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		4. BELA users questionnaire	Qualtrics questionnaire via email	SPSS analysis (Independent samples t-tests)
		5. ASC feedback survey	Survey Monkey via email	NVivo thematic analysis
		6. Uptake of support	Internal reports	Excel analysis
		7. Writing development gains	BELA and BELA 2 data via learning management systems	Excel analysis (comparative analysis)
		8. Student outcomes of those who do not take up support	Internal reports and learning management system data	SPSS analysis (correlational analysis)

	Research question	Data type	Data collection method	Data analysis method
Predicting academic performance	RQ2 What factors predict academic performance of university students?	1. Student academic performance data (i.e., GPA after two semesters) and student background data	Learning management system data and internal reports	Binary logistic regression

This page has been intentionally left blank.