



**THE DEVELOPMENT AND ARGUMENT-BASED  
VALIDATION OF AN ANALYTIC RUBRIC FOR  
L2 WRITING ELECTRONIC PORTFOLIOS**

**BY**

**SHEILA PARVEEN LALLMAMODE**

**A thesis submitted in fulfilment of the requirement  
for the degree of Doctor of Philosophy**

**Kulliyah of Islamic Revealed Knowledge and Human  
Sciences  
International Islamic University  
Malaysia**

**JULY 2012**

## ABSTRACT

Electronic portfolios are used in many institutions worldwide. Despite the numerous benefits, few studies have investigated the usefulness of the ePortfolio in the writing classroom, in general, and looked into providing reliable and valid scoring rubrics to assess L2 writing ePortfolios, in particular. The purpose of this study is to build a validity argument for WASPER, an analytic rubric developed for L2 writing ePortfolios, using Bachman and Palmer's (1996) test qualities and Bachman's (2005) Assessment Use Argument as a basis. Validity is established by evaluating the empirical data that serve as the backing for the inferences of the interpretive argument, in light of evidentiary reasoning. The five claims of the argument-based approach to validity are: (1) WASPER is a reliable tool to assess L2 academic writing ePortfolios; (2) WASPER is relevant to the construct being measured; (3) WASPER corresponds to a more authentic way of assessing academic writing for L2 students in the TLU domain; (4) WASPER has the potential to positively impact on the teaching and learning of L2 academic writing; and, (5) WASPER's development was practical, and it is a practical scale to use to assess writing ePortfolios. The L2 writing ePortfolio assessment was implemented for one semester with B.Ed. students, at the Institute of Education, International Islamic University Malaysia. Thirty-eight students participated in this study, along with thirteen raters. The mixed method approach, and the Process Model (Chatterji, 2003) for the design and validation of rating scales, were both used to develop and validate WASPER. Both qualitative and quantitative data were collected to serve as evidence to back the warrants or reject the rebuttals in the validity argument. The Many-facet Rasch Measurement (MFRM) was used to evaluate the quantitative data. Anonymous feedback forms, standard open-ended interviews and classroom observation notes were used to elicit the raters' and students' perceptions of the usefulness of the scale and the ePortfolio assessment in general. The results indicated that the five-point scoring rubric WASPER is a reliable and valid instrument for the purpose of assessing L2 writing ePortfolios. The FACETS analysis showed that the WASPER rating scale categories were all ordered and functioned appropriately. A reliability of 0.81 with a 19% error variance indicated that the examinees were well differentiated in three levels of ability. However, it was found that the criterion *Ease of Navigation* should be removed from WASPER as it is related and has an impact on the assessment of the other criteria. Evidence was also found that WASPER corresponds to an authentic way of assessing L2 writing and has a positive impact on students and raters. The limitations and implications of the use of WASPER, as well as suggestions for refinement in future rounds of validations are discussed.

## خلاصة البحث

لا يزال العديد من المؤسسات في مختلف أنحاء العالم يترددون في الاعتماد على استخدام البورت فوليو الإلكتروني (ePortfolios)، رغم أنه يحتوي على فوائد عديدة. ولوحظ أن الدراسات التي بحثت في حدوده قليلة خاصة فيما يتعلق بدروس الكتابة بصفة عامة، وكذلك في تقديم نماذج معتمدة لتقييم البورت فوليو الإلكتروني للكتابة في اللغة الثانية. لذلك تهدف هذه الدراسة إلى إيجاد برهان معتمد لاستخدام (WASPER) وهو نموذج تحليلي مطور باستخدام مميزات الاختبار لباتشمان و بالمر (1996) واستخدام جودة التقييم لباتشمان (2005). وقد كانت المواصفات الخمس للمنهجية القائمة على البرهان كما يلي : (1) WASPER هو أداة موثوقة لتقييم البورت فوليو الإلكتروني للكتابة الأكاديمية في اللغة الثانية. (2) WASPER وثيق الصلة ببناء ما يجري قياسه. (3) WASPER يناظر بطريقة أكثر واقعية لتقويم الكتابة الأكاديمية لدى الطلبة في مجال استخدام اللغة الثانية المستهدفة (TLU). (4) WASPER لديه القدرة على التأثير الإيجابي على تعليم وتعلم الكتابة الأكاديمية في اللغة الثانية. (5) WASPER مقياس عملي لتقويم البورت فوليو الإلكتروني للكتابة. كانت فترة التقويم فصلاً دراسياً واحداً لطلبة البكالوريوس في معهد التربية بالجامعة الإسلامية العالمية ماليزيا وكان عددهم ثمانية وثلاثون طالبا وطالبة وعدد المقومين ثلاثة عشر. وقاموا بتجميع كل من البيانات النوعية والكمية، ثم استخدم نموذج راش (Rasch) متعدد الأوجه لتقويم البيانات الكمية. واستخدمت نماذج التغذية الراجعة والمقابلات المفتوحة والملاحظات الصفيحة على تصورات المقومين والطلبة. أشارت النتائج إلى أن المواصفات الخمس لـ (WASPER) هي أداة موثوقة ومعتمدة لتقييم البورت فوليو الإلكتروني للكتابة في اللغة الثانية. وأظهر تحليل نموذج راش متعدد الأوجه لتقويم بأن فئات مقياس WASPER مرتبة وتعمل بشكل مناسب. وأشارت قيمة الثبات (0.81) مع قيمة فرق الخطأ (19%) إلى أن قدرة الطلبة كانت متباينة بشكل جيد في مستوياتهم الثلاثة وأخيراً قدمت مقترحات لتحسين عمليات التقويم القادمة.

## APPROVAL PAGE

The thesis of Sheila Parveen Lallmamode has been approved by the following:

---

Nurairhan Mat Daud  
Supervisor

---

Noor Lide Abu Kassim  
Co-Supervisor

---

Joharry Othman  
Internal Examiner

---

Carol Siri Johnson  
External Examiner

---

Ibrahim Zein  
Chairman

## DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Sheila Parveen Lallmamode

Signature .....

Date .....

INTERNATIONAL ISLAMIC UNIVERSITY  
MALAYSIA

**DECLARATION OF COPYRIGHT AND  
AFFIRMATION OF FAIR USE OF UNPUBLISHED  
RESEARCH**

Copyright@2012 by International Islamic University Malaysia. All rights reserved.

**THE DEVELOPMENT AND ARGUMENT-BASED VALIDATION OF AN  
ANALYTIC RUBRIC FOR L2 WRITING ELECTRONIC PORTFOLIOS**

I hereby affirm that The International Islamic University Malaysia (IIUM) hold all rights in the copyright of this work and henceforth and reproduction or use in any form or by means whatsoever is prohibited without the written consent of IIUM. No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form of by means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder.

Affirmed by Sheila Parveen Lallmamode

.....  
Signature

.....  
Date

This study is dedicated to:

My late father, Abdool Hakim Lallmamode (*Rahimahullah*), who was one of the most  
inspirational figures in my life.

He wiped my tears when I failed and instilled in me the notion of helping others and  
also, most importantly,

to strive for excellence in whatever I am doing.

## ACKNOWLEDGEMENTS

First and foremost, all praise belongs to Allah, the One and Only God who has no son or daughters. I thank Allah for allowing me to complete this thesis to the best of my ability and ask Allah to help me make a difference for the better in the academic field that I have chosen as my profession. Prayers and peace be upon His Prophet Mohammad, the last messenger for mankind and jinns.

Looking back at my life and at the many pleasurable years spent at the International Islamic University Malaysia for my academic degrees, I am grateful to having had the opportunity to study there and to meeting some hard-working faculty members who provided me with the intellectual stimuli to help me make a difference.

Reaching the completion of thesis writing would not be possible without the support, guidance, and understanding of several key people.

To the one who inspired me for my MA thesis and who later became my PhD main supervisor, I have to thank Prof. Dr. Nuraihan Mat Daud. May Allah reward you. You did not give up on me, especially when I had to return to Saudi Arabia to be with my family. I would also like to thank you for showing faith in me that helped me reach higher than I thought I could. No mere words can express my gratitude.

It requires only a glance at the quantitative data analysis to grasp the magnitude of my debt to Assoc. Prof. Dr. Noor Lide Abu Kassim, my second supervisor. She is my mentor in RASCH measurement and Language Testing. I thank you for your unwavering faith in me and also for your patience, kindness, wisdom, humor and friendship. You will continue to inspire me in all my future endeavours, Insha Allah.

I would also like to thank two people whom I have not physically met but who assisted me when I felt that Rasch measurement was above what I could digest: Mike Linacre and Carol Myford. They always responded to emails promptly, and did not hesitate to send me whatever journal papers or notes I needed.

This research was also only possible through the involvement of a large cast of B.Ed., MA and PhD students at the International Islamic University Malaysia. They gave their time and energy to make it possible, and produced wonderful ePortfolios. To them I express my sincere thanks for their assistance. I am also thankful to all the wonderful raters who participated in this study despite their busy schedules.

I cannot fail to mention my gratitude to colleagues and friends, both in Malaysia and Saudi Arabia, who made the research and writing up of this thesis a smoother process. Among them are Bashir Dedans, Dr. Kamal Badrasawi, Farida Osman, Dr. Ainol Zubairi, Dr. Fayez Al-Ghamdi, Ebtisam Abdul Halim, Sarah Adham and Dr. Normala Othman.

My special thanks go to Abu Huzaifah Ayman Al-Dakkak and to his dear, loving family. They corrected me in the best possible ways when I erred, and gave me the best source of oxygen for me to carry on, which is adherence to the Sunnah, away



from all forms of extremisms and deviations. I have never ever seen people with such a sense of hospitality, and being with them was like being home while being away from home... May Allah reward them.

I am also grateful for the financial support I received for this study from the Research Centre at the International Islamic University Malaysia.

My deepest gratitude go to my immediate family: my loving mother, Farida, who provided genuine support and sincere advice to me in all of my entire life; and, to my husband, Abdur Rahman (Raphael) Pallaud, for his patience and unwavering strong support, in all ways possible. I would also like to thank my sister and brother, as well as my nephews and nieces for their support when I am 'home' in Mauritius. I also have to thank my in-laws in France for all their kind support and understanding when I had to sacrifice family visits during the years I spent on this thesis.

Finally, I should not forget to mention my sweet cats 'Zoulou', 'Satte' and 'Toutounette', who accompanied me during the writing up process and who spent countless days under my desk, patiently waiting for me to complete this thesis. Writing can be a lonely task, and the quiet friendship of pets is a blessing. Their little paws should also be stamped on this thesis.

## TABLE OF CONTENTS

Abstract.....	ii
Abstract in Arabic.....	iii
Approval Page.....	iv
Declaration Page.....	v
Copyright Page.....	vi
Dedication.....	vii
Acknowledgements.....	viii
List of Tables.....	xviii
List of Figures.....	xxi
<b>CHAPTER 1: INTRODUCTION .....</b>	<b>1</b>
1.1 Statement of the Problem .....	9
1.1.1 Studies in L2 Writing ePortfolio .....	10
1.1.2 Assessing both the Process and Products in ePortfolios .....	12
1.1.3 Evaluating the Usefulness of the ePortfolio for L2 students .....	13
1.2 An Argument-based Approach to Validation.....	16
1.3 Objectives of the Study .....	20
1.4 Significance of the Study.....	23
1.5 Background of the Study.....	24
1.6 Methodology.....	25
1.7 Organisation of Chapters.....	28
1.8 Definitions of terms.....	29
<b>CHAPTER 2: LITERATURE REVIEW.....</b>	<b>31</b>
2.1 The Electronic Portfolio.....	31
2.1.1 Fundamental Characteristics of Portfolios/ePortfolios.....	32
2.1.2 Theories/Models of Writing and their Roles in Portfolios.....	33
2.1.3 Constructivist and Positivist Portfolios.....	37
2.1.4 Benefits and Challenges of Writing Electronic Portfolios.....	39
2.1.5 Recommendations before the Implementation of the ePortfolio.....	42
2.1.5.1 Steps to Develop Electronic Portfolios.....	43
2.1.5.2 Customized Software and Generic Tools.....	44
2.1.5.3 What should be Include in the Writing ePortfolio? ...	45
2.1.5.4 How Many Pieces of Writing Tasks should be Included? .....	48
2.1.5.5 Who Decides What to Include in the ePortfolio? .....	49
2.1.6 Studies in Writing Portfolios/ePortfolios.....	52
2.2 Issues Surrounding L2 Academic Writing .....	66
2.2.1 The L2 Writing Construct.....	66
2.2.2 Features of Academic Writing.....	68
2.2.3 Characteristics of L2 Writing in the Academic Context.....	69

2.2.4	Process Writing.....	72
2.2.4.1	Stages in Process Writing.....	74
2.2.4.2	Process Writing and L2 Students' ePortfolios.....	77
2.2.5	Self- and Peer-feedback in L2 writing.....	80
2.2.6	The Computer in the L2 Writing Classroom.....	84
2.3	Developing a Tool For Assessing L2 Academic Writing .....	88
2.3.1	Factors in Rater-mediated Writing Assessments.....	89
2.3.2	The Administration Procedures.....	94
2.3.3	Electronic Writing Portfolio Scoring.....	95
2.3.4	The Construction of Rating Scales .....	97
2.3.4.1	Data-based Scale Construction.....	97
2.3.4.2	Empirically Underlying Scale Construction.....	98
2.3.4.3	Intuition Based Construction of Rating Scales.....	100
2.3.4.4	Theory-based Method of Scale Development.....	100
2.3.5	The Main Types of Scoring Rubrics Used in Assessing Writing	103
2.3.5.1	Primary and Multiple Trait Scoring.....	103
2.3.5.2	Holistic Scoring.....	105
2.3.5.3	Analytic Scoring.....	109
2.3.5.4	Combining Both Analytic and Holistic Scales in a Rubric.....	114
2.3.6	Issues Related to the Development of Criteria in a Rubric.....	121
2.3.6.1	Assessing Both the Writing Process and Product .....	121
2.3.6.2	Assessor-, User- and Constructor-Oriented Rubrics...	125
2.3.6.3	Criteria for Writing ePortfolio Rubrics .....	125
2.3.6.4	The Number of Acceptable Criteria in a Rubric.....	128
2.3.6.5	Context-specific Dimensions for Developing Criteria..	129
2.3.7	Developing Descriptors and Categories.....	131
2.3.7.1	The Number of Acceptable Categories.....	132
2.3.7.2	Labelling and Formulation of the Descriptors.....	133
2.3.7.3	The Native Speaker as a Sign of Achievement.....	135
2.3.8	Assessing Digital Artifacts in Electronic Portfolios.....	137
2.3.9	Issues Related with Raters.....	139
2.3.9.1	Rater Characteristics.....	139
2.3.9.2	Providing Handbook to Raters .....	143
2.3.9.3	The Standardisation Session.....	143
2.3.9.4	Use of Anchors in Clarifying Descriptors.....	146
2.3.9.5	The Rating Session.....	146
2.4	Validity & Validation of Assessment Tools.....	147
2.4.1	Validity.....	147
2.4.2	Frameworks & Validation of Rating Scales.....	153
2.4.2.1	Mixed Research as a Tool for Developing Quantitative Instruments.....	155
2.4.2.2	The Socio-cognitive Validity Framework.....	158
2.4.2.3	The Process Model.....	159
2.4.2.4	Bachman's Assessment Use Argument.....	163
2.4.3	Reliability.....	168
2.4.4	Utility or Practicality tests.....	170
2.4.5	Backwash or Impact.....	171

2.5	Validation According to Rasch Measurement.....	171
2.5.1	Multi-faceted Rasch Measurement.....	173
2.5.2	Rasch Measurement and Construct Validity Measures.....	177
2.5.3	Studies in Validation of Rating Scales with FACETS.....	181
2.6	Chapter Summary.....	190
<b>CHAPTER 3: METHODOLOGY.....</b>		<b>191</b>
3.1	Research design.....	191
3.1.1	Mixed Method.....	191
3.1.2	Building a Validity Argument for WASPER.....	193
3.1.3	The Process Model.....	197
3.2	Phase I: The Project.....	200
3.2.1	The Purpose.....	200
3.2.2	The Constructs and Observable Indicators.....	200
3.2.3	The Academic Writing ePortfolio.....	201
3.2.3.1	The ePortfolio Content.....	201
3.2.3.2	Instructions Provided to Students.....	205
3.2.4	Study Timeline.....	209
3.3	Phase II: The Assessment Specifications.....	210
3.3.1	The Candidates.....	210
3.3.2	Results of the Demographics, Writing & Familiarity with Computers Questionnaire.....	212
3.3.2.1	Summary of the Results.....	213
3.3.3	The Raters.....	214
3.3.4	The Rating Manual.....	215
3.3.5	Selection of the Anchors.....	216
3.3.6	The Standardisation Sessions.....	216
3.3.7	The Roles of the Researcher and the Instructor.....	220
3.4	Phase III: Developing The Scoring Rubric.....	220
3.4.1	The Selection of Criteria.....	221
3.4.2	The Dimensions of the Rubric.....	224
3.4.3	The Descriptors of the Rubric.....	225
3.5	Phase IV: The Content Validation and Empirical Evaluations of WASPER.....	226
3.5.1	The Content Validation of the Assessment Tool.....	226
3.5.2	The Empirical Evaluations of the WASPER Rubric.....	227
3.6	Stage 1 : The Pilot Study.....	227
3.6.1	The Participants.....	228
3.6.2	The Debriefing Session.....	229
3.7	Stage 2 : Empirical Evaluation – Quantitative Data Analysis.....	232
3.7.1	Global Model Fit.....	234
3.7.2	Reliability Measures.....	235
3.7.2.1	Reliability Separation Indexes.....	236
3.7.2.2	Examinee Separation Ratio.....	237
3.7.2.3	Rater Point-biserial Coefficient (Inter-rater reliability).....	237
3.7.2.4	The Percentage of Exact Agreement.....	238
3.7.2.5	Variation in Ratings.....	238

3.7.3	Rater Effects.....	239
3.7.3.1	Leniency/severity.....	241
3.7.3.2	Central tendency.....	242
3.7.3.3	Randomness Effect.....	242
3.7.3.4	Halo Effect.....	243
3.7.3.5	Differential Leniency / Severity Effect.....	243
3.7.4	Bias Analysis.....	244
3.7.5	Steps to Improve the Functionality of a Rating Scale.....	245
3.7.6	Construct Validity Measures.....	248
3.7.6.1	Category Frequencies.....	249
3.7.6.2	Average Observed Measures.....	249
3.7.6.3	Category Thresholds.....	250
3.7.6.4	Mean-Square Outfit Statistics.....	250
3.7.6.5	Category Probability Curves.....	251
3.7.6.7	Meaningful Variable and Uni-dimensionality.....	252
3.7.7	Authenticity Measures.....	252
3.8	Stage 2 : Empirical Evaluation – Qualitative Data Collection and Analysis.....	253
3.8.1	Standardised Open-ended Interviews with the Raters.....	254
3.8.2	Group Interviews with the Students.....	255
3.8.3	Anonymous Feedback Form.....	256
3.8.4	Classroom Observation Notes.....	257
3.9	Chapter Summary.....	258
 <b>CHAPTER 4: RESULTS OF THE QUANTITATIVE DATA ANALYSIS</b>		<b>259</b>
4.1	General Overview.....	259
4.1.1	Data Connectivity.....	260
4.1.2	The Variable Map.....	262
4.1.3	Global Model Fit.....	267
4.2	The Reliability Measures of WASPER.....	268
4.2.1	Examinee Discrimination Measures.....	268
4.2.2	Homogeneity of the Raters.....	271
4.2.3	Rater Separation and Reliability.....	273
4.2.4	Rater Agreement.....	275
4.2.5	Consistency in the Use of the Rating Scale.....	276
4.2.6	Ratings Assigned by Particular Raters.....	282
4.2.7	Point-biserial Correlation of the Raters.....	284
4.2.8	Rater Effects.....	285
4.2.8.1	The Professional Milieux of the Raters.....	286
4.2.9	Bias Analysis.....	288
4.2.10	Steps Taken to Improve the Reliability Measure.....	290
4.2.10.1	Collapsing Categories in WASPER.....	291
4.2.10.2	Removing Misfitting ePortfolios or Items.....	291
4.3	The Construct Validity Measures in WASPER.....	293
4.3.1	Rating Scale Effectiveness.....	293
4.3.1.1	The Category Frequencies.....	294
4.3.1.2	The Average Observed Measures.....	294
4.3.1.3	The Category Threshold .....	295

4.3.1.4	The Mean-Square Outfit Statistics.....	295
4.3.1.5	The Category Probability Curves.....	296
4.3.2	The Spread of ePortfolio Marks.....	298
4.3.3	Infit and Outfit Statistics of the Criteria in WASPER.....	298
4.3.4	Separation Indices.....	299
4.3.5	Point-biserial Correlation for each Criterion.....	301
4.3.6	Category Functioning of each Criterion.....	302
4.3.6.1	Substance & Content.....	304
4.3.6.2	Organisation.....	305
4.3.6.3	Language Use.....	306
4.3.6.4	Textual Attribution.....	308
4.3.6.5	Writing Process.....	309
4.3.6.6	EP Design.....	311
4.3.6.7	Use of Multimedia.....	313
4.3.6.8	Ease of Navigation.....	314
4.3.6.9	Holistic.....	316
4.3.6.10	Summary of the Category Functioning of each Criterion.....	317
4.4	Authenticity of the WASPER Rubric.....	318
4.4.1	Comparability of the Analytic and Holistic Scoring Scales.....	318
4.4.1.1	The Analytic Scales.....	319
4.4.1.2	The Holistic Scale.....	320
4.4.1.3	Correlation Analysis Between Analytic and Holistic Scores.....	324
4.4.2	Correlation Between the writing and ePortfolio scales.....	326
4.5	Conclusion.....	327
<b>CHAPTER 5: RESULTS OF THE QUALITATIVE DATA ANALYSIS...</b>		<b>328</b>
5.1	Factors Affecting Reliability.....	330
5.1.1	Students' Perceptions of the Usefulness of ePortfolio Assessment.....	331
5.1.1.1	Improvement of Writing and Digital Writing Skills....	333
5.1.1.2	The Self- and Peer-evaluations.....	333
5.1.1.3	Insufficient Time for Completion of Task.....	334
5.1.1.4	Insufficient Support from the Researcher and Course Instructor.....	336
5.1.1.5	The Correction Codes and Symbols.....	338
5.1.1.6	Limited Resources Available.....	338
5.1.2	Raters' Perceptions of the Usefulness of the ePortfolio Assessment.....	339
5.1.2.1	Benefits and Drawbacks of Using the ePortfolio.....	339
5.1.2.2	The Standardisation Sessions.....	344
5.1.2.3	The Anchors.....	345
5.1.2.4	The Rater Handbook.....	346
5.1.2.5	Length of Time Allocated to Mark the ePortfolios.....	348
5.1.2.6	Marking at Home.....	348
5.1.2.7	Effect of the Conditions of Raters on the Rating Process.....	349

5.1.3	Rater Effects and Biases.....	349
5.1.3.1	Distinctive Characteristics of some of the Raters.....	350
5.1.3.2	Strategies Used by Raters.....	356
5.1.4	Summary.....	357
5.2	Factors Affecting Construct Validity.....	358
5.2.1	Perceptions of the Effectiveness of WASPER.....	358
5.2.1.1	Students' Perceptions of the Effectiveness of WASPER.....	358
5.2.1.2	Raters' Perceptions of the Effectiveness of WASPER.....	359
5.2.2	Comprehensiveness of WASPER.....	359
5.2.3	Effects of Construct-irrelevant Variance and Construct Underrepresentation.....	360
5.2.3.1	Substance & Content.....	360
5.2.3.2	Organisation.....	361
5.2.3.3	Language Use.....	363
5.2.3.4	Textual Attribution.....	365
5.2.3.5	Writing Process.....	366
5.2.3.6	EP Design.....	369
5.2.3.7	Use of Multimedia.....	370
5.2.3.8	Ease of Navigation.....	372
5.2.4	Summary.....	372
5.3	Factors Affecting Authenticity.....	373
5.3.1	The Criteria and Scoring Method in WASPER Correspond to a More Authentic Way of Assessing Writing for L2 students....	374
5.3.1.1	Writing Process Criterion Adding to Authenticity in WASPER.....	374
5.3.1.2	Usefulness of the Holistic Scale in WASPER.....	375
5.3.2	The EP has been Tailored to Correspond to a More Authentic Way of Assessing Writing for L2 Students.....	376
5.3.2.1	Effectiveness of the PowerPoint as a Digital Authoring Environment.....	376
5.3.2.2	Usefulness of the Self- and Peer-evaluations.....	379
5.3.3	Summary.....	380
5.4	Factors Affecting Impact.....	381
5.4.1	Impact on the Students.....	381
5.4.2	Impact on the Raters.....	384
5.4.3	Summary.....	386
5.5	Conclusion.....	386
<b>CHAPTER 6: DISCUSSION AND CONCLUSION.....</b>		<b>388</b>
6.1	The Argument-based Validity.....	388
6.1.1	Reliability.....	389
6.1.1.1	Warrant 1: Raters rate reliably when using the scale...	389
6.1.1.2	Rebuttal 1: Rater effects and bias items .....	391
6.1.2	Summary.....	394
6.1.3	Construct Validity.....	394
6.1.3.1	Warrant 2: The criteria successfully discriminate between examinees.....	394

6.1.3.2	Warrant 3: The scale provides empirical and substantive evidence .....	396
6.1.3.3	Rebuttal 2: There are effects of construct-irrelevant variance and construct under-representation.....	398
6.1.4	Summary.....	400
6.1.5	Authenticity.....	401
6.1.5.1	Warrant 4: The criteria and scoring method in WASPER correspond to a more authentic way.....	401
6.1.5.2	WARRANT 5: The EP has been tailored to correspond to a more authentic way .....	403
6.1.6	Summary.....	404
6.1.7	Impact.....	405
6.1.7.1	Warrant 6: Provisions are made for involving students directly, or for collecting and utilising feedback .....	405
6.1.7.2	Warrant 7: The impact on using the ePortfolio assessment and WASPER was positive .....	406
6.1.8	Summary.....	407
6.1.9	Practicality.....	407
6.1.9.1	Warrant 8: The Scale Development is Practical.....	408
6.1.9.2	Warrant 9: The Scale is Practical to Use.....	409
6.1.10	Summary.....	410
6.1.11	Conclusion.....	410
6.2	Limitations of the Study.....	412
6.2.1	Limitations in the Development Stage of WASPER and the Implementation of the ePortfolio Assessment Programme.....	412
6.2.2	Limitations in the Argument-based Validation of WASPER.....	414
6.3	Implications.....	415
6.4	Suggestions for Future Research.....	417
6.5	Final Remarks.....	420
<b>BIBLIOGRAPHY.....</b>		<b>421</b>
<b>APPENDICES.....</b>		<b>446</b>
Appendix A	B.Ed. Course Outline.....	446
Appendix B	Course Learning Outcome Matrix.....	448
Appendix C	Assessment Systems and ePortfolios .....	449
Appendix D	Rubric for Electronic Portfolio, University of Wisconsin.....	450
Appendix E	Electronic Portfolio Template .....	452
Appendix F	WASPER Rubric for Peer-Feedback [2nd Draft].....	454
Appendix G	WASPER Rubric (without descriptors).....	456
Appendix H	Information about each Artifact Required in the ePortfolio.....	458
Appendix I	Digital Error Correction Code & Symbols.....	460
Appendix J	Letter Seeking Permission from the Institute of Education.....	461
Appendix K	Letter of Consent for Participants.....	463
Appendix L	Demographics, Writing and Familiarity with Computers Questionnaire.....	464



Appendix M	Results of the Demographics, Writing and Familiarity with Computers Questionnaire.....	467
Appendix N	Letter of Invitation to Lecturers and Instructors.....	472
Appendix O	Handbook for Raters.....	474
Appendix P	Final Version of the WASPER Rubric.....	485
Appendix Q	Letter to Content Reviewers.....	487
Appendix R	Detailed Version of WASPER Rubric Used in the Pilot Study	488
Appendix S	Interview Questions to Raters.....	491
Appendix T	Interview Questions to Students.....	492
Appendix U	Anonymous Feedback Form for the Students.....	494

## LIST OF TABLES

<u>Table No.</u>		<u>Page No.</u>
1.1	Trends in Assessment	1
1.2	List of Warrants and Rebuttals for Each Test Quality	22
2.1	Criteria for Technical Communication Rubric	58
2.2	Six Features of Academic Writing	68
2.3	Characteristics of L1 and L2 English Language Student Writers	72
2.4	Differences between Process Writing and Product Writing	76
2.5	Lists of Facets in a Writing Assessment	92
2.6	A Checklist for Designing Portfolio-based Assessments	95
2.7	Advantages and Disadvantages of the Holistic and Analytic Scoring	116
2.8	A Comparison of Holistic and Analytic scales	117
2.9	Definitions of the Four Dimensions	131
2.10	Various Facets that Influence Raters	140
2.11	Past and Current Conceptions of Validation	151
2.12	Actions to be undertaken at each phase in the Process Model	162
3.1	Flow Chart of the Process Model as Applied to the Design and Validation of WASPER	199
3.2	Constructs and Observable Indicators in this Study	201
3.3	Detailed Table of Contents of the L2 Academic Writing ePortfolio	203
3.4	Demographic Data of the Raters	214
3.5	The Timetable for the Five Standardisations Sessions	217
3.6	Criteria and Descriptors of WASPER	223
4.1	The Rotating Test Book Judging Plan	261

4.2	Descriptive Statistics of the Ratings	263
4.3	Examinee Discrimination Measures	270
4.4	Distribution of Analytic Ratings Across Raters	272
4.5	Summary of Rater Separation & Reliability Measures	275
4.6	Raters with Misfitting Infit and Outfit MnSq	277
4.7	Unexpected Responses of the Thirteen Raters	281
4.8	Ratings Assigned by Selected Raters	282
4.9	Determining Rater Effects	286
4.10	Severity Measures of Raters According to Kulliyah / CELPAD	287
4.11	Unexpected Ratings From Rater 11	292
4.12	Unexpected Responses for the Criterion Ease of Navigation	292
4.13	Summary of the Six Statistical Indicators	298
4.14	Criteria Discrimination Measures	301
4.15	Summary of the Five Statistical Indicators of Rating Scale Effectiveness for each Scale in the WASPER Rubric	318
4.16	Summary of the Five Statistical Indicators of Rating Scale Effectiveness	324
5.1	Qualitative Research Questions Based on the Test Qualities	329
5.2	Factors Affecting the Reliability of WASPER	331
5.3	Summary of Results about Self- and Peer-evaluations	334
5.4	Strategies Used by Raters to Mark the ePortfolios	356
6.1	Summary of Qualitative and Quantitative Evidence with regard to Reliability	394
6.2	Summary of Qualitative and Quantitative Evidence with regard to Construct Validity	401
6.3	Summary of Qualitative and Quantitative Evidence with regard to Authenticity	405
6.4	Summary of Qualitative and Quantitative Evidence with regard to Impact	407

6.5	Summary of Qualitative Evidence with regard to Practicality	410
6.6	Summary of all the Qualitative and Quantitative Evidence for the Assessment Use Argument	411

## LIST OF FIGURES

<u>Figure No.</u>		<u>Page No.</u>
1.1	Learning Landscape Framework	9
1.2	Structure of Bachman's Assessment Use Argument	19
2.1	A Summary of the Principal Orientations to L2 Writing Teaching	35
2.2	Portfolio Characteristics and Theories of Writing	37
2.3	Conceptual Model – The Intersection of Assessment and Instruction	41
2.4	Dimensions and Elements of the Holistic Portfolio Scoring Rubric	51
2.5	Extract from the Portfolio Evaluation Form	54
2.6	Analytic Rubric for the Assessment of Online Portfolios	59
2.7	A Recursive Model of Writing	75
2.8	PowerPoint Slides of a Research Electronic Portfolio	79
2.9	A Peer-review introductory Sheet	82
2.10	Cognitive-affective and Behavioural Responses of L2 Students	86
2.11	Factors in Writing Assessment	90
2.12	Interaction of Facets in Performance Assessment	91
2.13	A Conceptual–psychometric Framework of Factors Relevant in Rater-mediated Performance Assessments	93
2.14	A Hypothetical Four Category Writing Scale	98
2.15	The EBB Scale	99
2.16	A Multi-trait Scoring Rubric for Factual Recount	105
2.17	Guide for Preparing Holistic Scoring Rubrics	108
2.18	Model Showing Factors that Affect Holistic Scoring	108
2.19	The ESL Composition Profile	110

2.20	Analytic Rubric for Assessing Electronic Portfolios	112
2.21	Research Paper Rubric for Grade 10 History	114
2.22	A Six-point Scale Analytic and Holistic Score Sheet for Portfolio	119
2.23	Student Feedback Form - ESL Services Courses	124
2.24	Criteria Used in the Scoring of ESL writing by Commonly Known Standardised Tests	127
2.25	Synthesis of Writing Constructs Elements	127
2.26	Dimensions for Assessing Portfolios	130
2.27	The Different Types of Formulating Descriptors for Rating Scales	134
2.28	Electronic Portfolio Rubric	138
2.29	Facets of Validity Matrix	152
2.30	The Interpretive Argument	153
2.31	Mixed Research-based Meta-framework	156
2.32	Crossover Analysis Strategies by Validity	157
2.33	The Socio-cognitive Validity Framework	159
2.34	A Flow Chart of the Process Model	161
2.35	Excerpt of the Characteristics of Test Usefulness for a Test	166
2.36	Summary of Warrants for the Validation of the New DELNA Scale	168
2.37	Classical Test Theory and Raters' Interpretation of Rating Scales	174
2.38	Validity Evidence Using the Rasch Models	180
3.1	Facets of Rating Scale Validity	194
3.2	A Validity-argument for the Usefulness and Efficacy of WASPER	196
3.3	Schematic Representation of the Phases of the Research Design	198
3.4	The Constructs, Observable Indicators and Criteria in this Study	200
3.5	Snapshot of the Class Website for B. Ed Students	206
3.6	Rating Scale Continuum of WASPER for the Students	208
3.7	Study Timeline	210

3.8	Dimensions for the WASPER ePortfolio Assessment Programme	225
3.9	Statistical Indicators from FACETS for Five Common Rater Effects	244
3.10	Average Measure Differences for Two Rating Scales	246
3.11	Qualitative Data to be Used as Evidence	254
4.1	Calibrations of the Raters, Examinees (ePortfolios) and the WASPER scales on a Logistic Scale	264
4.2	FACETS Data Summary	267
4.3	Examinee Measurement Report	269
4.4	Raters Measurement Report	271
4.5	Partial-credit Model: Raters 7 and 11 Individual Rating Patterns	279
4.6	Category Statistics Report of Raters 7 and 11	280
4.7	Rater Partial Credit Model Rating Scale Structures	284
4.8	Standard Errors Around Rater Severity Measures	288
4.9	Bias Interactions between Raters and Items	288
4.10	Rater-by-item Bias Interactions of Some Raters	290
4.11	Measurement Results for the Category Statistics	293
4.12	Category Probability Curves Showing Category Response Peaks	296
4.13	Criteria Measurement Report	300
4.14	Partial Credit Facets Map	303
4.15	Measurement Results for the Category Statistics for <i>Substance &amp; Content</i>	304
4.16	Category Probability Curves for <i>Substance &amp; Content</i>	305
4.17	Measurement Results for the Category Statistics for <i>Organisation</i>	305
4.18	Category Probability Curves for <i>Organisation</i>	306
4.19	Measurement Results for the Category Statistics for <i>Language Use</i>	307
4.20	Category Probability Curves for <i>Language Use</i>	308

4.21	Measurement Results for the Category Statistics for <i>Textual Attribution</i>	308
4.22	Category Probability Curves for <i>Textual Attribution</i>	309
4.23	Measurement Results for the Category Statistics for <i>Writing Process</i>	310
4.24	Category Probability Curves for <i>Writing Process</i>	311
4.25	Measurement Results for the Category Statistics for <i>EP Design</i>	311
4.26	Category Probability Curves for <i>EP Design</i>	312
4.27	Measurement Results for the Category Statistics for <i>Use of Multimedia</i>	313
4.28	Category Probability Curves for <i>Use of Multimedia</i>	313
4.29	Measurement Results for the Category Statistics for <i>Ease of Navigation</i>	315
4.30	Category Probability Curves for <i>Ease of Navigation</i>	315
4.31	Measurement Results for the Category Statistics for <i>Holistic</i>	316
4.32	Category Probability Curves for <i>Holistic</i>	317
4.33	Measurement Results for the Category Statistics for Only the Analytic Scales	319
4.34	Category Probability Curves for Only the Analytic Scales	320
4.35	Facets Map for Only the Holistic Scale	321
4.36	Category Use and Step Difficulty for the Holistic Scale	322
4.37	Probability Curves for the Holistic Scale	323
4.38	Scatterplot of the Holistic and Analytic Ability Measures	325
4.39	Plot of Examinee Measures Based on the Writing and ePortfolio Scales	326
5.1	Self- and Peer-evaluation Exercises in the Computer Lab at IIUM	383