Supporting English Language Learners with an Adaptive Mobile Application

by

Carrie Anne Demmans Epp

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

© Copyright by Carrie Anne Demmans Epp 2016
Supporting English Language Learners with an Adaptive Mobile Application

Carrie Anne Demmans Epp

Doctor of Philosophy

Graduate Department of Computer Science

University of Toronto

2016

Abstract

English language learners (ELL) have dedicated considerable time and effort to the development of their language proficiency. This has included the use of a variety of mobile assisted language learning (MALL) tools that are either unproven or that have undergone limited evaluations of their effectiveness. The majority of these evaluations have been performed with beginner foreign-language learners at the post-secondary level. Moreover, dedicated MALL tools rarely support the learner’s ability to communicate in English. I propose and demonstrate the feasibility of an adaptive MALL approach that aims to scaffold ELL vocabulary and communication needs. This scaffolding recommends learning materials to ELLs by employing the ecological approach to dynamically reason over logs of learner interactions with a MALL tool.
The highly personalized approach to supporting learners that is operationalized through this tool was developed following user-centered design principles. The development of the learning content generation and recommendation mechanisms that are included as part of this approach to supporting English language learners was validated through two studies. An additional exploratory evaluation of this adaptive approach to supporting ELL communication and learning activities was performed before evaluating its influence on ELL vocabulary knowledge, communication, and affect through two studies.

These studies considered the effectiveness of the proposed MALL approach from multiple perspectives. The first took place in a Japanese high school and focused on the relationship between student vocabulary knowledge and system usage. The second involved advanced English language learners and took place in the greater Toronto area. This study aimed to determine the relationships among system usage, user communicative success, and user affect. The work presented in this thesis shows that the use of the proposed approach can support ELL communication, vocabulary development, and affect. The evaluation of this approach allowed the creation of models that predict learning outcomes based on learners’ MALL usage and knowledge. Combining the results of these studies with those of the formative evaluations, indicates that a mobile tool that employs the ecological approach to learner modeling can support the learning activities, vocabulary learning outcomes, affect, and communication of English language learners.
Acknowledgements

I owe a number of people for their guidance, support, and mentorship. This list includes the faculty who were on my committee and others who were willing to provide me with an outside perspective on my work, my career, and my academic program. Gord McCalla, Julita Vassileva, Daniel Wigdor, Susan Bull, and Khai Truong were kind enough to question me when I needed questioning and provide another perspective from which I could view the world and my place within it.

I am grateful to the members of DGP, ARIES, the AIED community, and TAGlab for their feedback and the help that they provided with piloting and participant recruitment. I would especially like to thank the exceptional undergraduate students and staff who helped with developing VocabNomad: Garry Ing, Stephen Tsourounis, Justin Djordjevic, and Chris Arnold.

I am grateful to the faculty, research groups, and programs who made me feel welcome and who enabled my professional development even though they had no obligation to do so. These people and groups include Julie Kerekes, Rhonda McEwen, the PeppeR research team, the Centre for Educational Research on Languages and Literacies (CERLL), and the Geva lab.

I would like to thank my committee members. Esther Geva was kind enough to give me the guidance and mentorship that I needed with respect to research, language learning, and assessment. Ravin Balakrishnan challenged me and helped me to consider options that I had not realized were available, and Cosmin Munteanu provided encouragement and support. My advisor, Ron Baecker, provided me with numerous professional development opportunities and advice. He also enabled me to gain valuable experience that will help me going forward.

Beyond those who helped me with my research, I must thank those who supported me emotionally. Most important of these would be my spouse and the University of Toronto synchronized swimming team who were there for me when everything appeared to be going wrong. I also owe thanks to several colleagues and co-authors for their support and guidance. This list includes Christopher Brooks, Daita Watanabe, Gina Park, Christopher Plumb, Chris Harwood, Matthew Johnson, Krystle Phiranggee, Heather Birch, and Mayya Sharipova.
I would not have completed the work contained within this thesis were it not for the above people or the kindness of the teachers and students at the schools that enabled my recruitment of participants and the deployment of MALL applications.
Table of Contents

Chapter 1 – Introduction .............................................................................................................. 1

1.1 The Status of English and English Language Learners ................................................. 1

1.2 The Status of Mobile Assisted Language Learning ...................................................... 3

1.3 VocabNomad Design and Development ........................................................................ 5

1.4 VocabNomad Evaluations ............................................................................................... 7

1.5 Thesis Organization ......................................................................................................... 11

Chapter 2 – Background Literature ....................................................................................... 12

2.1 Language Learning .......................................................................................................... 12

  2.1.1 The Role of Language Exposure or Input ................................................................. 13
      2.1.1.1 Direct Study ...................................................................................................... 14
      2.1.1.2 Learning from Context ....................................................................................... 15

  2.1.2 The Role of Language Production ............................................................................. 16

  2.1.3 The Role of Affect ..................................................................................................... 17

  2.1.4 Learner Modeling and Adaptation ............................................................................ 19

2.2 Computer Assisted Language Learning .......................................................................... 20

  2.2.1 The Role of Language Exposure or Input ................................................................. 21
      2.2.1.1 Direct Study ...................................................................................................... 22
      2.2.1.2 Learning from Context ....................................................................................... 22

  2.2.2 The Role of Language Production ............................................................................. 22

  2.2.3 The Role of Affect ..................................................................................................... 23

  2.2.4 Learner Modeling and Adaptation ............................................................................ 24

2.3 Mobile Assisted Language Learning ............................................................................... 25

  2.3.1 The Role of Language Exposure or Input ................................................................. 28
      2.3.1.1 Direct Study ...................................................................................................... 28
4.2 User Interface Validation .......................................................... 63

4.3 On-Demand Vocabulary Support: Development and Validation Studies .................. 66

4.3.1 Generating Collections of Vocabulary On-Demand ...................................66
    4.3.1.1 Instruments and Procedures ..................................................67
    4.3.1.2 Participants ........................................................................67
    4.3.1.3 Results .............................................................................68
    4.3.1.4 Discussion ........................................................................70
    4.3.1.5 Conclusion .........................................................................72

4.3.2 Image Retrieval and Matching .............................................................72
    4.3.2.1 Instruments and Procedures ..................................................72
    4.3.2.2 Participants ........................................................................73
    4.3.2.3 Results .............................................................................74
    4.3.2.4 Discussion and Conclusion ....................................................75

4.4 Conclusion: Feature Design and Validation ............................................... 76

Chapter 5 - VocabNomad: System Description .............................................. 77

5.1 Expected Uses ..................................................................................... 77
    5.1.1 User Personae .........................................................................77
    5.1.2 Personae Use of VocabNomad ....................................................78

5.2 User Interface ...................................................................................... 82
    5.2.1 Mobile Interface ......................................................................82
    5.2.2 Web Interface ..........................................................................83

5.3 System Architecture ............................................................................ 86
    5.3.1 Domain Model ..........................................................................87
        5.3.1.1 Developer-Generated Content ..........................................90
        5.3.1.2 System-Generated Content .............................................90
        5.3.1.3 User-Generated Content .................................................91
    5.3.2 Model of the World ...................................................................92

5.3.3 Learner Model ..................................................................................92
5.3.3.1 Data ................................................................................................................................................. 92
5.3.3.2 Modeling Process ............................................................................................................................... 95
5.3.3.3 Adaptive Theory ............................................................................................................................... 96

5.4 Conclusion .................................................................................................................................................. 99

Chapter 6 - VocabNomad: System Validation Study ..................................................................................... 100

6.1 Instruments and Procedures ..................................................................................................................... 101

6.2 Participants .............................................................................................................................................. 101

6.3 Results ...................................................................................................................................................... 102
  6.3.1 Language-Learning Strategies .......................................................................................................... 102
  6.3.2 Technology Use ................................................................................................................................... 103
  6.3.3 VocabNomad Use ................................................................................................................................ 103

6.4 Discussion ............................................................................................................................................... 105

6.5 Conclusion .............................................................................................................................................. 106

Chapter 7 - Vocabulary Study in an English as a Foreign Language Class .................................................. 109

7.1 Methodology ........................................................................................................................................... 112
  7.1.1 Researcher Introduction and Integration ............................................................................................ 113
  7.1.2 Tool Introduction, Explanation, and Training .................................................................................... 114
  7.1.3 Consent and Remuneration .............................................................................................................. 114
  7.1.4 Data Collection .................................................................................................................................. 115
    7.1.4.1 User Observation ......................................................................................................................... 116
    7.1.4.2 Demographics Information ........................................................................................................... 116
    7.1.4.3 Learner Self-reports of Language Proficiency ............................................................................. 117
    7.1.4.4 Assessments of Language Knowledge ......................................................................................... 118

7.2 Research Site ......................................................................................................................................... 120
  7.2.1 Course Organization: EI2 ................................................................................................................ 121
  7.2.2 Course Organization: Other Courses ............................................................................................... 122
  7.2.3 Learning Activities: EI2 .................................................................................................................. 122
7.2.4 Learning Activities: Other Courses

7.3 Participant Demographics

7.3.1 EI2 Teachers

7.3.2 Other Teachers

7.3.3 Students

7.4 Results

7.4.1 Student Reactions to the Devices

7.4.2 Student Reactions to VocabNomad

7.4.3 VocabNomad Usage

7.4.3.1 Logged Usage

7.4.3.2 Observed Usage

7.4.4 Morphological Knowledge

7.4.5 PPVT-4

7.4.6 Final Exam Vocabulary Scores

7.4.7 Self-Assessed Language Ability

7.4.8 The Relationship Between Presentation Groups and Measures

7.4.9 English Language Abilities

7.4.10 The Relationship Between Usage and Language

7.5 Barriers and Challenges Faced

7.6 Successes

7.7 Opportunities

7.8 Discussion

7.8.1 Measures

7.8.2 Research Questions

7.8.3 Student Responses
7.8.4 Implications for Educators ........................................................................................................ 154

7.9 Stakeholder Reports .......................................................................................................................... 155

7.9.1 Student Reports ............................................................................................................................ 156

7.9.2 School Report ................................................................................................................................ 156

7.10 Limitations ......................................................................................................................................... 157

7.11 Conclusion ......................................................................................................................................... 157

Chapter 8 - Affect and Communication Study in an English Language Environment .......... 159

8.1 Methodology ...................................................................................................................................... 160

8.1.1 Recruitment and Consent ............................................................................................................... 161

8.1.2 Data Collection and Measures ....................................................................................................... 161

8.1.2.1 Affect ......................................................................................................................................... 161

8.1.2.2 Communication: Number of Attempts and Perceived Success Rates ........................................ 163

8.1.2.3 Perceived Competence Scale (PCS) ............................................................................................ 164

8.1.2.4 MyLog – An Experience Sampling Application ............................................................................. 164

8.1.2.5 Demographic Information ........................................................................................................ 167

8.1.2.6 Interviews .................................................................................................................................. 167

8.1.2.7 Language Knowledge .................................................................................................................. 167

8.1.3 Participants ..................................................................................................................................... 168

8.2 Results .............................................................................................................................................. 169

8.2.1 Technology Use ............................................................................................................................... 169

8.2.2 VocabNomad Use ............................................................................................................................ 171

8.2.3 MyLog Use ..................................................................................................................................... 172

8.2.4 Affect .............................................................................................................................................. 174

8.2.4.1 Momentary Affect ....................................................................................................................... 174

8.2.4.1 Daily Affect .................................................................................................................................. 175

8.2.4.2 Relationship Between Affect and Application Usage ................................................................. 176

8.2.5 Communication: Number of Attempts and Perceived Success Rates ........................................ 178

8.2.5.1 Perceived Momentary Communicative Success .............................................................................. 178
8.2.5.2 Perceived Daily Communicative Success ................................................................. 179
8.2.5.3 Number of Communication Attempts ................................................................. 180
8.2.5.4 Relationship Between Communication and Application Usage .......................... 181

8.2.6 Perceived Competence Scale (PCS) ....................................................................... 181

8.2.7 Language Knowledge .............................................................................................. 182
  8.2.7.1 Morphological Knowledge .................................................................................. 182
  8.2.7.2 Vocabulary Knowledge (PPVT-4) ...................................................................... 182
  8.2.7.3 Vocabulary Knowledge (Adaptive and Personalized Testing) ......................... 183
  8.2.7.4 Learner Self-Assessments ................................................................................... 184

8.3 Discussion .................................................................................................................... 186

8.3.1 Participant Communication (RQ6) ........................................................................ 186
8.3.2 Participant Affect (RQ7) ......................................................................................... 187

8.3.3 Language Knowledge ............................................................................................... 188
8.3.4 Measures .................................................................................................................. 189

8.3.5 Implications for Language Learning Technology .................................................... 190

8.4 Limitations .................................................................................................................. 191

8.5 Conclusion ................................................................................................................... 191

Chapter 9 – Discussion, Future Work, and Conclusion ..................................................... 195

9.1 Research Questions and Hypotheses .......................................................................... 196
  9.1.1 RQ1: Extent of Support for Vocabulary Learning and Communication .............. 197
  9.1.2 RQ2: Integration into Language Learning and Communication Practices ............ 200
  9.1.3 RQ3: Learner Vocabulary Knowledge, Affect, Communicative Success, and System Usage ... 204

9.2 Contributions .............................................................................................................. 205

9.3 Future Work and Potential Expansions ..................................................................... 207
  9.3.1 Educational Opportunities ..................................................................................... 207
  9.3.2 MALL Development Opportunities .................................................................... 209
9.3.2.1 Adaptive Features and System Recommendations ................................................................. 209
9.3.2.2 Scaffolding ............................................................................................................................. 211
9.3.2.3 Feedback .............................................................................................................................. 212
9.3.2.4 Socio-Collaborative Support ............................................................................................... 213

9.3.3 Research Opportunities .......................................................................................................... 214

9.4 Summary ..................................................................................................................................... 216

References ......................................................................................................................................... 221
List of Tables

Table 1 Studies of system use and the measures that were employed.......................... 8

Table 2 Participant demographics – MyVoice proof-of-concept study.......................... 39

Table 3 Participant Technology Use – MyVoice Proof-of-Concept Study........................ 45

Table 4 A summary of the changes that were made to the basic system design as a result of the proof-of-concept study, its associated design recommendations (DR), and the extent to which the design recommendation is supported (ES). .................................................................................................................. 62

Table 5 Some of the stages in the design and refinement of the web-based user-interface. ...... 64

Table 6 Some of the stages in the design and refinement of the android client user-interface. ... 65

Table 7 The algorithm-corpus pairs that were evaluated.................................................. 67

Table 8 Examples of discourse responses that could still be understood, even if words had not been provided (missing words) by the algorithm-generated vocabulary. The actual response is the message that participants could communicate using the algorithm-generated vocabulary lists, the intended message shows how the participant had wanted to communicate that message. ..... 68

Table 9 Mean and standard deviation of response lengths in number of words.................. 68

Table 10 The average number of missing words and T-test results comparing the number of missing words from the combined vocabularies to that of the base vocabulary ($M = 5.07$, $SD = 2.35$). .......................................................................................................................... 70

Table 11 The average size of each vocabulary (in number of words) and proportion of each vocabulary that was used across all themes and locations (the base vocabulary was static). .... 70

Table 12 The image corpora that were evaluated. .............................................................. 73

Table 13 T-Test results comparing Corpus A to Corpus B for participant ratings of the appropriateness of an image for representing the meaning of a word. .......................... 75
Table 14 A subset of the different database entities and their role within the system’s domain model

Table 15 The learner’s characteristics portion of the learner model. These data describe the user’s more stable characteristics and are spread throughout the users, UserRole, UserCourses, activities, location, and Time tables in the database. They are presented jointly to facilitate comprehension of the model attributes rather than to provide details about the model’s implementation

Table 16 The episodic history portion of the learner model

Table 17 A listing and description of the event types that are logged by the web interface.

Table 18 A listing and description of the event types that are logged by the android interface.

Table 19 Example adaptive tests that are based on the frequency of use of a word within English (English Freq. Based) or the frequency with which participants interacted with the vocabulary item (Usage Based). ESM01 and ESM02 are the participants who completed these tests.

Table 20 Participant demographics – VocabNomad validation study

Table 21 Participant technology use to support their English – VocabNomad validation study.

Table 22 The data collection schedule for the Japanese high school study.

Table 23 Student application use (number of logged events) by study phase.

Table 24 Selected usage statistics by quartile.

Table 25 The usage patterns for each of the k-means identified clusters.

Table 26 The number of students assigned to each usage cluster when K = 3 and K = 4.

Table 27 The centroid values for each of the features (unitized event counts per week).

Table 28 Student scores (%) on morphological measures.
Table 29 PPVT-4 reliability statistics by administration.......................................................... 137

Table 30 PPVT-4 percentage scores by administration......................................................... 137

Table 31 Distribution of vocabulary scores from the final exam. .......................................... 139

Table 32 Shapiro-Wilk test p-values for participant self-assessments of their Japanese-language (J) abilities...................................................................................................................... 139

Table 33 Student self-assessed Japanese-language (J) proficiency........................................ 140

Table 34 Shapiro-Wilk test results for participant self-assessments of their English-language (E) abilities........................................................................................................................ 140

Table 35 Student self-assessed English-language (E) proficiency.......................................... 141

Table 36 The number of students assigned to each language-ability cluster when K = 3 and K = 4................................................................................................................................................. 143

Table 37 The confusion matrices for the W-J48 Decision trees.............................................. 145

Table 38 The data collection schedule for the communication and affect study. Black circles indicate when a measure was taken in relation to the study stages. ................................................. 167

Table 39 Participant demographic and remuneration information – Communication and Affect Study.................................................................................................................................................. 168

Table 40 Selected VocabNomad usage statistics (event counts). ........................................... 173

Table 41 Momentary affect scores (%) by phase................................................................. 174

Table 42 Daily affect scores (%) by phase. ........................................................................... 175

Table 43 Momentary communication success scores (%) by phase.................................... 178

Table 44 Daily communication success scores (%) by phase............................................. 179

Table 45 Daily communication attempts (number) by phase............................................. 181
Table 46 PCS scores (%) by administration. ................................................................. 181
Table 47 Participant scores (%) on morphological measures. ................................. 182
Table 48 PPVT-4 reliability statistics by administration ........................................... 182
Table 49 PPVT-4 scores (%) by administration ....................................................... 183
Table 50 Shapiro-Wilk test $p$-values for the adaptive and personalized testing by administration. ........................................................................................................ 183
Table 51 Adaptive and personalized testing scores (%) by administration ................ 184
Table 52 $P$-values for the relationships between the adaptive and personalized tests .......... 184
Table 53 Student self-assessed L1 (varied) and English (L2) proficiency ..................... 185
Table 54 Shapiro-Wilk test $p$-values for participant self-assessments of their L1 (varied) and English (L2) abilities ................................................................. 185
List of Figures

Figure 1 Countries of origin for those who migrate to Canada (mid-2013 estimates). ¹ 2

Figure 2 Some of the screens from VocabNomad’s mobile client: content recommendation (left), the results of a search (centre), and the detailed content of a vocabulary item (right). ................. 6

Figure 3 An overview of the types of measures that were used for the summative evaluations of VocabNomad as a whole. Purple squares indicate dependent variables and ovals indicate independent variables: aqua ovals were directly measured but pink ovals could not be directly measured when the exposure happened outside of the learning environment. ............................... 9

Figure 4 The two-dimensional view of affect. This is a modified version of the one presented in Russell (1980, p. 1164). .................................................................................................................. 18

Figure 5 MyVoice's web interface (left) and mobile interface (right). .............................................. 35

Figure 6 Shared writing space from participant O........................................................................... 37

Figure 7 Participant use of MyVoice to support their learning and communication goals. ........ 52

Figure 8 The layers of the evaluation framework and their interactions. ................................. 59

Figure 9 A screen capture of the interface for the image study questionnaire. ......................... 74

Figure 10 The number of ratings performed by participants. ..................................................... 75

Figure 11 A storyboard showing the screens that Jota would see were he recording a newly encountered word........................................................................................................ 79

Figure 12 A storyboard showing the screens that Tara would encounter were she trying to learn how to say honeydew.............................................................................................. 80

Figure 13 A storyboard showing the screens that Koji would encounter when studying or editing a word.................................................................................................................... 81

Figure 14 The vocabulary list (left), a single word (centre), and its edit screen (right). ............ 82
Figure 15 The search functionality. Both word-based (left) and tag-based (centre) options are shown. The results for a tag or category-based search are shown on the right................................. 83

Figure 16 The Vocabulary list view of the web interface following a search for milk. ............... 84

Figure 17 The page where users can import collections of vocabulary items. ............................. 84

Figure 18 The test page. This test has been configured to include translation and sentence production tasks. ......................................................................................................................... 85

Figure 19 The page where users can see which items they have deleted and retrieve them. ....... 85

Figure 20 The page where users can add or edit vocabulary items. ................................. 86

Figure 21 VocabNomad’s architecture as seen through the lens of the layered approach. ......... 87

Figure 22 A learning object reuse example for the smallest of the learning objects (images). .... 88

Figure 23 The pseudo code for the on-demand vocabulary generation feature......................... 91

Figure 24 A state diagram that details the recommendation conditions for synonyms. $VC_i$ is the view count for the specified vocabulary entry, $S_i$ is the visibility of synonyms for vocabulary entry i, and $S_{si}$ is the synonym visibility of i’s synonyms from the perspective of vocabulary item i only. $S_{si}$ could be true if another one of its synonyms (e.g., j), is known (i.e, $VC_j \geq 4$). ............ 96

Figure 25 Synonym recommendation propagation for SUV, with lines between entities indicating synonymous relationships. The system has inferred that the learner knows SUV (the star). Its first-degree connections (the hexagons: Sport Utility Vehicle and Vehicle) also display their synonyms, but second degree or higher connections do not display their synonyms (the circles: truck and lorry). ............................................................................................................................ 97

Figure 26 Participant uses of VocabNomad to support their learning and communication goals. ............................................................................................................................................... 104

Figure 27 The layout of a student’s desk prior to the first class (bottom). The classroom’s layout from the front (top-left) and back (top-right)......................................................... 117
Figure 28 An example of how a PPVT-4 stimulus page might look for the word sun. This example is fabricated to indicate the nature of the test while preserving its integrity. .............................. 119

Figure 29 Examples of student on-task behavior: translating a word that Nihon_S25 needed for his presentation (centre), and a customized home screen where the background image is related to Nihon_S13’s presentation topic (left). The camera icon indicates that Nihon_S13 also customized his device by integrating his Gmail and Picasa accounts. An image that Nihon_S20 downloaded (right); it related to his presentation topic. ................................................................. 129

Figure 30 Nihon_S37’s response for one of the items from a VocabNomad-related worksheet. He translated handout as print out (purinto). ................................................................. 134

Figure 31 PPVT-4 scores (M and SD) by group and administration (left). PPVT-4 score gains (M and SD) by group and phase (right). .................................................................................. 138

Figure 32 Participant self-assessed English-language ability (M and SD) by group and administration (left). English-language ability score gains (M and SD) by group and phase (right). .................................................................................. 141

Figure 33 Decision rules predicting student final exam performance. .............................................. 144

Figure 34 W-J48 Decision trees that predict a student’s PPVT-4 score gain category. The tree for K = 3 is on the left, and K = 4 is on the right. ................................................................. 146

Figure 35 The affect measures as they appear in the experience sampling application: SAM with momentary instructions (left; black) and the I-PANAS-SF (all others). The momentary (centre-left; black), daily reflective (centre-right; yellow), and weekly reflective (right; green) time instructions are shown for the I-PANAS-SF. ................................................................. 162

Figure 36 The communication measures as they appear in the experience sampling application: both the momentary (left; blue) and the reflective daily (right; yellow) versions are shown. 163

Figure 37 The PCS as it appears in the experience sampling application: both the learning (left; My Log and Page6) and the communication (right; Page7 and Page8) versions are shown. 164
Figure 38 MyLog: application signaling and postponing interface (left; black) and time instructions (all others): momentary affect (left-centre; black), daily (right-centre; yellow), and weekly (right; green).................................................................................................................. 165

Figure 39 MyLog: additional contextual questions related to location and social context (left; black), affect (centre-right; green), and participant activities (right; yellow). The participant activities screen also provides an example of input validation.......................................................... 166

Figure 40 Advanced ELL technology use to support language-learning activities.................... 170

Figure 41 The purposes for which participants used VocabNomad.............................................. 172

Figure 42 Participant tracking information for daily affect measures; dashed lines indicate the boundaries between phases.............................................................................................................................................. 177

Figure 43 Participant tracking information for daily communicative success; dashed lines indicate the boundaries between phases. ........................................................................................................................................... 180
List of Appendices

Appendix A – Ethics Approval .................................................................................................................. 246
Appendix B – Demographics Forms ....................................................................................................... 251
Appendix C – Interview Guides ............................................................................................................. 254
Appendix D – Image Attributions & Permissions .................................................................................. 259
Appendix E – Algorithms ....................................................................................................................... 260
Appendix F – Discourse Completion Tasks ........................................................................................... 263
Appendix G – VocabNomad Entity-Relationship Diagram ...................................................................... 265
Appendix H – Adaptive Testing Development ....................................................................................... 267
Appendix I – Testing Instruments ........................................................................................................ 269
Appendix J – Japan: Course Worksheets & Reports ............................................................................... 275
Appendix K – Japan: Student Data ........................................................................................................ 288
Appendix L – Publications Related to this Thesis .................................................................................. 292
### List of Definitions and Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation</td>
<td>The amount of energy that is being experienced.</td>
</tr>
<tr>
<td>Arousal</td>
<td>See Activation</td>
</tr>
<tr>
<td>Blended Learning</td>
<td>Learning that occurs in multiple environments. Typically, this involves combining the classroom environment with a technology-based environment.</td>
</tr>
<tr>
<td>Bootstrapping</td>
<td>The cognitive process that enables fast mapping; it allows us to create cognitive placeholders for symbols so that we can later assign them meaning and refine those meanings based on how we encounter the word being used in context. The initial understanding of the word’s meaning may be arrived at through the fast mapping process. Subsequent refinements to our understanding of a word happen as we gain continued exposure to that symbol.</td>
</tr>
<tr>
<td>CALL</td>
<td>See Computer Assisted Language Learning (CALL)</td>
</tr>
<tr>
<td>Ceiling Effects</td>
<td>When an independent variable reaches the level where its variance is no longer measured or it no longer influences the dependent variable.</td>
</tr>
<tr>
<td>Computer Assisted Language Learning (CALL)</td>
<td>The use of computer applications to help people with language teaching or learning.</td>
</tr>
<tr>
<td>Cronbach’s Alpha</td>
<td>This is a lower-bound reliability coefficient. It represents the internal consistency of a set of items as measured by the correlation among them.</td>
</tr>
<tr>
<td>Discovery with Models</td>
<td>When a model is used as an input to another modeling process.</td>
</tr>
<tr>
<td>Ecological Approach</td>
<td>A form of learner modeling that supports pre-determined and emergent modeling needs through a dynamic modeling process rather than a static model. This approach requires that learner information that remains true for extended periods of time (e.g., his or her address or mother tongue) is stored independently of the learner’s interaction</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>history</td>
<td>Models of learner knowledge are generated when a need for that model arises. These dynamic models are computed based on the information about a learner’s interactions with learning materials.</td>
</tr>
<tr>
<td>E-learning</td>
<td>Any form of teaching or learning that is supported by electronic tools.</td>
</tr>
<tr>
<td>ELL</td>
<td>See English Language Learner (ELL)</td>
</tr>
<tr>
<td>English Language Learner (ELL)</td>
<td>A person who is attempting to learn English.</td>
</tr>
<tr>
<td>Extended Mapping</td>
<td>The cognitive process by which someone achieves a more nuanced understanding of a symbol (e.g., word or phrase) based on repeated exposure to that symbol being used in and across contexts.</td>
</tr>
<tr>
<td>Fast Mapping</td>
<td>The cognitive process by which someone learns a new piece of information (e.g., word or phrase) based on a single exposure to that piece of information.</td>
</tr>
<tr>
<td>First Language (L1)</td>
<td>The first language that a person learns.</td>
</tr>
<tr>
<td>Foreign Language</td>
<td>A language that is not the person's mother tongue and is being used outside of an environment in which it is the dominant language. For example, studying Cree at the University of Toronto.</td>
</tr>
<tr>
<td>Guttmann Split-Half Reliability</td>
<td>A measure of the relationship between two parts of the same scale or test. This is a lower-bound form of split-half reliability.</td>
</tr>
<tr>
<td>Improvement Rate</td>
<td>The percentage of improvement between two phases of a study. IRD Difference (IRD) is one of the options for determining effect size in single-subject designs.</td>
</tr>
<tr>
<td>Intelligent Tutoring System (ITS)</td>
<td>Any computer program or system that provides adaptive instruction, feedback, or support to learners.</td>
</tr>
<tr>
<td>IRD</td>
<td>See Improvement Rate Difference (IRD)</td>
</tr>
<tr>
<td>ITS</td>
<td>See Intelligent Tutoring System (ITS)</td>
</tr>
</tbody>
</table>
| Layered Evaluation                        | A framework for evaluating adaptive systems. This framework decomposes the system's adaptive functionality and allows individual
components to be validated separately from the system as a whole.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>See First Language (L1)</td>
</tr>
<tr>
<td>L2</td>
<td>See Second Language (L2)</td>
</tr>
<tr>
<td>LCMS</td>
<td>See Learning Content Management System (LCMS)</td>
</tr>
<tr>
<td>Learner Model</td>
<td>A representation of the learner's knowledge, abilities, actions, beliefs, and motivations.</td>
</tr>
<tr>
<td>Learning Analytics</td>
<td>The collection, measurement, and analyses of data about learners and their learning contexts. This is usually performed with the intent of understanding the learner and his or her environment. Learning analytics are also meant to support the improvement of learning and the environments in which it occurs.</td>
</tr>
<tr>
<td>Learning Content Management System (LCMS)</td>
<td>A type of LMS that focuses on the development, management, and publishing of learning materials.</td>
</tr>
<tr>
<td>Learning Management System (LMS)</td>
<td>A computer program that supports the planning, delivery, and management of learning events.</td>
</tr>
<tr>
<td>Learning Object</td>
<td>A collection of material that can be used to facilitate learning. The collection can contain a single item or many related items. Learning objects are self-contained and reusable.</td>
</tr>
<tr>
<td>Learning Objective</td>
<td>A performance-based description of what a student should be able to do following a learning unit.</td>
</tr>
<tr>
<td>LMS</td>
<td>See Learning Management System (LMS)</td>
</tr>
<tr>
<td>MALL</td>
<td>See Mobile Assisted Language Learning (MALL)</td>
</tr>
<tr>
<td>Mobile Assisted Language Learning (MALL)</td>
<td>The use of mobile devices and applications to support language learning.</td>
</tr>
<tr>
<td>Language</td>
<td>A cognitive process that is related to language learning. It involves the production and negotiation of meaningful language.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Morpho-orthographic Choice Task</td>
<td>A language assessment where people must select the correctly spelled word from a pair.</td>
</tr>
<tr>
<td>Morphological Structure Test</td>
<td>A language assessment where people must transform a word to fit within a provided sentence.</td>
</tr>
<tr>
<td>Mother Tongue</td>
<td>The first language that a person learns. It is possible for someone to have multiple mother tongues.</td>
</tr>
<tr>
<td>OLM</td>
<td>See Open Learner Model (OLM)</td>
</tr>
<tr>
<td>Open Learner Model (OLM)</td>
<td>A learner model where the person can access at least some of the information that the model has about him or her. This information may be presented in an abstract form as an indirect measure of the user’s abilities.</td>
</tr>
<tr>
<td>PANAS</td>
<td>See Positive and Negative Affect Schedule (PANAS)</td>
</tr>
<tr>
<td>Positive and Negative Affect Schedule (PANAS)</td>
<td>A validated measure of affect that can be subdivided into measures of the person’s positive and negative affective states.</td>
</tr>
<tr>
<td>PPVT-4</td>
<td>A standardized test of receptive vocabulary knowledge</td>
</tr>
<tr>
<td>SAM</td>
<td>See Self Assessment Manikin (SAM)</td>
</tr>
<tr>
<td>Scaffold</td>
<td>Support given during the learning process to help learners achieve new levels of knowledge or skills that the learner could not have achieved independently.</td>
</tr>
<tr>
<td>Second Language (L2)</td>
<td>A second language is a language that is not the person's mother tongue and that is being used in that language's environment. For example, studying Cree on a Cree Nation.</td>
</tr>
<tr>
<td>Self Assessment Manikin (SAM)</td>
<td>A validated measure of user affect that employs visual representations of a person’s affective state.</td>
</tr>
<tr>
<td>Target Language</td>
<td>The language that a person is attempting to learn.</td>
</tr>
<tr>
<td>Technology Enhanced Learning (TEL)</td>
<td>Learning that is supported through the use of technology.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>TEL</td>
<td>See Technology Enhanced Learning (TEL)</td>
</tr>
<tr>
<td>Test-Retest Reliability</td>
<td>The correlation between two administrations of the same measure</td>
</tr>
<tr>
<td>Unbiased Alternate Forms Reliability</td>
<td>A measure of the equivalence between two versions of the same measure (e.g., two different test forms).</td>
</tr>
<tr>
<td>Valence</td>
<td>The amount of pleasure being experienced.</td>
</tr>
<tr>
<td>Zone of Proximal Development (ZPD)</td>
<td>The difference between what a learner can do unaided and aided.</td>
</tr>
<tr>
<td>ZPD</td>
<td>See Zone of Proximal Development (ZPD)</td>
</tr>
</tbody>
</table>
Chapter 1 – Introduction

English is recognized as a world language (Jenkins, 2009). Many people need to find ways to learn English since it is not their primary language and it is needed to conduct business across national borders. Language classes, a typical approach to language learning, do not always meet the needs of learners so learners are turning to other resources such as computer programs and applications on their mobile phones. Unfortunately, these applications fall short of providing English language learners (ELL) with the language-learning and communication support that they need because they rely on simple drill and practice or grammar-translation approaches. The use and effectiveness of these computational learning tools is also not well understood, especially with respect to their ability to support the vocabulary acquisition and communication of different types of ELLs across contexts. To address these shortcomings, we designed and built a mobile application that employs a different approach to supporting ELL needs. This new adaptive approach, which is exemplified through an application called VocabNomad, aims to support ELL communication and vocabulary learning needs. In what follows, I will discuss the status of English language learners and their current use of mobile applications for supporting their needs before briefly describing the design and evaluation of this new approach to using mobile tools to support the communication and learning needs of ELLs.

1.1 The Status of English and English Language Learners

There are economic incentives and benefits to mastering the language that is used in one’s surrounding environment. For immigrants in Canada, these benefits include a potential 10-12 percent increase in earnings (Saiz & Zoido, 2005). Since 20.7 percent of the 2013 Canadian population were migrants (MPI Data Hub, 2014b), up to one-fifth of the population could be economically disadvantaged when compared to those who have mastered English.

We know that nearly 600,000 of those included in the 2011 Canadian Census do not know either English or French (Statistics Canada, 2012) and the situation is worse in the United States where 50 percent of the foreign-born population does not speak English well (Gambino, Acosta, & Greco, 2014). Many of those who have moved to Canada from abroad come from non-English speaking countries (see Figure 1) (MPI Data Hub, 2014a), indicating that their language
proficiency may be below that necessary for accessing employment and social support (Gordon, 2004).

Even though English is an important skill to master, many people have difficulty achieving that mastery. People who speak English as a second, foreign, or subsequent language, hereafter called English language learners, often struggle with the various forms of English that are required in everyday life (Gordon, 2004). This difficulty is evidenced by the test scores of American ELLs in their final year of high school, where those who graduate receive lower scores than their English speaking peers on assessments of vocabulary knowledge (11.6 %), reading (11.0 %), and writing (19.0 %) (NAEP Data Explorer, 2014).

If we consider those who are living in environments that limit one’s exposure to English, we can easily believe that these ELLs face a greater learning challenge than those who live in English language environments where exposure to English and easy access to English-language materials is commonplace. The lack of exposure to understandable English along with the many other challenges that ELLs face independent of their environment indicates a need for more effective
communication support tools and learning strategies. ELLs need tools that support their vocabulary learning and communication practices across educational contexts.

1.2 The Status of Mobile Assisted Language Learning

Many researchers and language-teachers have suggested that mobile tools can be used to support both opportunistic learning outside of school (Ballance, 2013; Dearman & Truong, 2012; Edge, Searle, Chiu, Zhao, & Landay, 2011; Kukulska-Hulme & Shield, 2008) and classroom activities within schools (Ballance, 2013; Liu, 2009; Tan-Hsu Tan & Tsung-Yu Liu, 2004; Wong & Looi, 2010). However, the distinct needs of ELLs and the manner in which mobile tools can meet these needs are not yet widely understood.

Within formal educational contexts, ELLs may require additional support or access to different resources if they are to complete regular classroom activities (Graves, 2013). This differentiation can be the result of the highly variable needs of ELLs in foreign language contexts or the result of integrating ELLs into a class with their English as a first language (L1) peers. Mobile devices can provide this type of differentiated support without drawing special attention to those who are using the support tool. This personalized support is possible because of the computational power that modern smartphones possess and because of their wide adoption within the general populace (Kim-Rupnow & Burgstahler, 2004; Ludlow, 2001).

However, the use of mobile assisted language learning (MALL) tools is still predominantly a fringe activity (Burston, 2014b). We do not yet know how these tools affect the classroom when they are used over extended periods of time or when they are used as support tools rather than as the focus of the classroom learning activity. Most research into MALL has been design-based and has not considered the learning outcomes or other effects of MALL. The study of MALL use by privileged populations in homogeneous monolingual educational contexts is common, especially when measuring the effects of MALL or exploring its uses (Burston, 2014b; Kukulska-Hulme, 2013; Traxler, 2013): MALL use by undergraduate foreign language learners is better understood than its use by other populations because MALL use by immigrants, high school students, and adults who are not post-secondary students has received little attention.
We also do not know how effectively MALL tools support ELLs outside of the classroom. We have seen instances of MALL tools supporting short-term gains in vocabulary knowledge but have not seen evidence of their ability to support the communication of ELLs in everyday settings. Supporting ELL communication is necessary since members of this population are often isolated (Siegel, Martin, & Bruno, 2001) and sometimes rely on mobile translators or family and friends to interact with others in their environment (Gordon, 2004). However, the use of communication support tools has yet to be adopted by this community even though they have been employed to overcome the communication barriers that are faced by members of other populations (e.g., those who have had strokes) (McNaughton & Bryen, 2007).

In spite of the potential for mobile tools to support ELL communication and language learning, we have not seen many attempts at supporting both. Any tool that aims to support the user’s communication and his or her language learning must support vocabulary learning because vocabulary knowledge is an essential component of all of the macro-skills that make up language learning and usage. This is why I have set out to design, develop, and evaluate a MALL tool that can support ELL communication and vocabulary acquisition. Its evaluation takes place across educational contexts and focuses on the following primary research questions:

**RQ1.** To what extent can an adaptive mobile tool support both the vocabulary learning and communication of ELLs?

**RQ2.** How do users integrate an adaptive mobile tool into their language learning and communication practices?

**RQ3.** What is the relationship between the usage of an adaptive mobile tool and ELL vocabulary knowledge, communicative success, and affect?

Additional research questions are asked within individual chapters. These questions are more narrowly focused than RQ1, RQ2, and RQ3.

The investigations used to answer the above questions and evaluate the approach to providing MALL support that is operationalized through VocabNomad indicate that this approach can support the communication and learning activities of ELLs. These investigations also indicate
that this approach supports improvements in ELL affect and the development of their vocabulary knowledge.

1.3 VocabNomad Design and Development

Before these questions could be broken down into their constituent parts and answered, a tool that held the potential to support both the vocabulary learning and communication of ELLs had to be built. The development of this application started with the inspiration of a class of tools that support the communication of those who cannot communicate on their own (Demmans Epp, Campigotto, Levy, & Baecker, 2011). This resulted in the development of an adaptive MALL tool that is called VocabNomad (see Chapter 5 for details on the system architecture). Its development followed user-centered design practices with several forms of formative evaluation being employed (Chapter 4).

The first proof-of-concept study was conducted using the mobile communication support tool that inspired VocabNomad (Chapter 3). This study’s first aim was to verify that ELLs could use a communication support tool to scaffold their needs. The second purpose of this study was to inform the design of VocabNomad. This proof-of-concept study showed that these types of applications could in fact be used to support ELL communication and learning activities. Moreover, it resulted in the identification of several features and design elements that could be employed to better support ELLs. Chief among these features was the inclusion of on-demand support for emergent user needs.

VocabNomad was developed following this proof-of-concept study. Additional information was added and the content organization was changed. The user interface then underwent many rounds of refinement as the result of several types of formative assessment, including case studies and heuristic evaluation; Figure 2 shows some of the resulting screens. While developing the system and improving its user interface, we developed a feature that provides on-demand vocabulary support, based on a user-identified need. This feature has two parts that are used in combination to generate learning and support materials; the generation of support materials was integrated following the evaluation of these sub-components.
Figure 2 Some of the screens from VocabNomad’s mobile client: content recommendation (left), the results of a search (centre), and the detailed content of a vocabulary item (right).

The mechanisms that VocabNomad employs to generate these dynamic support materials from Internet-based corpora were validated for their ability to support communication across various contexts. The validation of the generated vocabulary lists was performed using a discourse completion study where 16 participants were required to respond to situations that commonly occur in different contexts. These situations included making requests, negotiating, and handling miscommunication (Demmans Epp, Djordjevic, Wu, Moffatt, & Baecker, 2012). The corpora that are used to supply visual representations of the meaning of the generated vocabulary were validated using a questionnaire that was conducted over the Internet ($N = 202$). The results from this questionnaire-based study indicated that the images that are retrieved from the evaluated corpora are sufficient for communicating the semantics of different types of English words. Details on the validation of this feature are provided in Chapter 4.

Since one of the objectives of building this new system was to provide ELLs with personalized support, adaptive features were added. The tool was implemented in a way that enables its later expansion, and detailed user tracking was added so that VocabNomad could use this information to infer the learner’s vocabulary knowledge. This detailed user tracking enables the use of a dynamic learner modeling technique, known as the ecological approach (McCalla, 2004), to perform these inferences and recommend additional learning materials. This approach to learner modeling divides the learner model into two components: the information that is true for an
extended period of time (i.e., the learner’s characteristics), such as his or her birth date, and a record of everything that the learner has done (i.e., the learner’s episodic history). Models of learner knowledge are then dynamically generated through a function that uses information from the learner’s characteristics and episodic history that are relevant to meeting the current modeling goal. This envisioning of the model as a process rather than an object enables it to meet new needs as they arise and distinguishes the ecological approach from other, more commonly used, learner modeling techniques (Desmarais & Baker, 2012; McCalla, 2004; VanLehn, 2006).

It may be easier to understand how users might interact with such a system or employ it to support their learning if we consider its use through two personae: Tara and Mohammad. Tara uses the system to review words and to help her recall certain words before she tries to communicate with people in the service industry. She may even use the application’s text to speech feature to communicate on her behalf or she may show one of the images to the person that she is speaking with to ensure clarity. Mohammad primarily uses VocabNomad to log words that are new to him or that he feels are important. He then studies the vocabulary items that are found within VocabNomad when he has time.

1.4 VocabNomad Evaluations

Once the basic user interface had been developed and the mechanisms for generating support materials had been integrated, a second deployment study was conducted with 8 ELLs. This study aimed to collect feedback about the design and usability of this new approach to supporting language learning, which is why more qualitative methods were used and the ELL’s self-reported language proficiency was measured (see Table 1). This deployment also aimed to collect information that would help answer the question about the extent to which an adaptive MALL tool could support both communication and vocabulary learning (RQ1). This deployment and the prior exploratory deployment showed that a system that employs image-word or phrase-word pairs, as is the case for VocabNomad, can scaffold the communication and vocabulary learning activities of ELLs (see Chapter 3 and Chapter 6 for details).

To understand how VocabNomad is perceived and used by ELLs, two additional studies were performed. Both studies used the same version of the software, but they focused on determining the relationship between different aspects of language learning and VocabNomad usage. The
first study (Chapter 7) focused primarily on vocabulary acquisition (RQ3) but also addressed questions about the tool’s integration into formal learning environments (RQ2). As a result of this study’s focus, measures of vocabulary knowledge and other factors that are known to influence learning were taken (see Table 1). The second study (Chapter 8) focused on the learner’s affective state and his or her ability to communicate with others using English (i.e., the learner’s communicative success). It aimed to address aspects of RQ1, RQ2, and RQ3. Figure 3 provides a visual overview of the types of measures that were considered during the two major summative evaluations of VocabNomad, and Table 1 summarizes the measures that were used for each study of the system as a whole. It also keys those studies to the thesis chapter in which they are detailed.

<table>
<thead>
<tr>
<th>Table 1 Studies of system use and the measures that were employed.</th>
<th>Formative Evaluations</th>
<th>Summative Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measures &amp; Contextual Information</strong></td>
<td><strong>MyVoice</strong></td>
<td><strong>VocabNomad</strong></td>
</tr>
<tr>
<td>Thesis Chapter</td>
<td>Chapter 3</td>
<td>Chapter 6</td>
</tr>
<tr>
<td>Chapter 7</td>
<td>Chapter 8</td>
<td></td>
</tr>
<tr>
<td><strong>Study Context</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational Environment</td>
<td>Informal</td>
<td>Informal</td>
</tr>
<tr>
<td>Participant Group</td>
<td>Recent Immigrants</td>
<td>Advanced ELLs</td>
</tr>
<tr>
<td>Location</td>
<td>Canada</td>
<td>Canada</td>
</tr>
<tr>
<td></td>
<td>Formal</td>
<td>Informal</td>
</tr>
<tr>
<td></td>
<td>High School Students</td>
<td>Advanced ELLs</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td>Canada</td>
</tr>
<tr>
<td><strong>Vocabulary Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardized (PPVT-4)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Adaptive Test</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Personalized Test</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td><strong>Other Language Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported Proficiency</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Morphological Production</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Morpho-orthographic Choice</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Affect</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-PANAS-SF</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SAM</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Self-reporting (interviews)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reporting (application)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Self-reporting (interviews)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>System Use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Logging</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Human observation</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Self-reporting (interviews)</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Figure 3 An overview of the types of measures that were used for the summative evaluations of VocabNomad as a whole. Purple squares indicate dependent variables and ovals indicate independent variables: aqua ovals were directly measured but pink ovals could not be directly measured when the exposure happened outside of the learning environment.

In addition to the goals of determining the relationship that VocabNomad usage has with vocabulary learning, ELL communication, and ELL affect, we set out to see if VocabNomad could be integrated into a classroom setting for an extended period (RQ2). The motivation for exploring VocabNomad’s integration into a classroom while determining its relationship to changes in ELL vocabulary knowledge comes from the lack of evidence that MALL tools have been successfully integrated into classrooms for extended periods (Burston, 2014a, 2014b).

The first summative assessment of the system as a whole is, therefore, the term-long integration of VocabNomad into a high school English as a foreign language class (Chapter 7). This study explores the relationship between VocabNomad usage and student vocabulary knowledge (RQ3). It also explores how to integrate this type of tool into a classroom setting where the tool’s use is not the primary focus of students’ learning activities (RQ1 and RQ2). This deployment study meets the call for longer-term investigations into the effects of MALL tools on the learning environment as well as on learning (Burston, 2014b).
This three-month long study was conducted in a Japanese high school using a methodology that is consistent with action-research (Avison, Baskerville, & Myers, 2001; Avison, Lau, Myers, & Nielsen, 1999): participating teachers were treated as collaborators and the school received a report detailing areas where improvements could be made. Students also received a report along with a workbook to help them improve their English. Various established measures of student knowledge of English morphology and vocabulary were taken, classroom activities were observed, student actions within the application were logged, and student work products were collected (see Table 1).

During this study, VocabNomad was used by 47 students. Improvements in vocabulary knowledge and perceived language proficiency were observed when VocabNomad was being used. The log data was combined with information about student language knowledge, and predictive models were built using a discovery with models approach. These models explain the relationship between various aspects of learner knowledge and system usage. In addition to highlighting some of the boundaries that prevent classroom adoption of MALL, the data from this study improve upon our understanding of how to effectively integrate these types of tools within classroom settings: students completed tasks and activities that were new to them and that their classroom teachers felt might be outside of their range of abilities.

The second summative evaluation of VocabNomad was the communication and affect study (Chapter 8). This was conducted in the greater Toronto area and consisted of a set of case studies that employed experience sampling within a single-subject with reversal design. The study objective was to determine how VocabNomad usage related to participant affective states and their ability to communicate successfully with others in their English language environment (RQ3). This focus meant that it was important to track learner actions within the system and to measure their affective state and perceived communicative success (Table 1). This study also explored how advanced ELLs, from varied backgrounds, integrated VocabNomad into their communication and learning practices (RQ2), which made the use of interviews and regular self-report data appropriate for capturing this data.

Data from the participant reports, qualitative interviews, assessments of participant knowledge, and application usage logs revealed that participants could use VocabNomad to study
vocabulary, support their written communication, and support their comprehension of oral language. Participants also experienced improvements in their affective state and communicative success when they had access to VocabNomad.

These two studies along with the formative evaluations of VocabNomad’s design show that this MALL approach can support vocabulary learning activities, oral communication, written communication, improvements in self-perceived language ability, and vocabulary learning. VocabNomad use also appears to have supported user communicative success and affect. These achievements provide evidence for the usefulness of employing the ecological approach to learner modeling that enabled the provisioning of personalized learning experiences within VocabNomad. Please see Chapter 6, Chapter 7, Chapter 8, and Chapter 9 for more nuanced discussions of the results.

1.5 Thesis Organization

This thesis is organized around the development, validation, and evaluation of VocabNomad. This is preceded by Chapter 2 which provides the necessary grounding in educational theories and previous computer and mobile assisted language learning. This discussion of prior work is followed by the evaluations that were used to inform VocabNomad’s design (Chapter 3 and Chapter 4) and a description of the resulting system (Chapter 5). I then detail the deployment study that was used to validate the system as a whole (Chapter 6). This final formative evaluation is followed by two summative evaluations. The first considers VocabNomad’s integration into a high school foreign-language educational setting (Chapter 7). The second explores VocabNomad’s use by advanced learners in an English-language dominant environment (Chapter 8). I then conclude with a discussion that ties these studies together and details the complicated relationships that exist between system usage and four key aspects of user communication and language learning: language knowledge, communicative success, language learning activities, and affect (Chapter 9).
Chapter 2 – Background Literature

To understand the current state of mobile assisted language learning, we must first consider the fields that contributed to its development. This includes language learning and learning technologies. We start with a discussion of language learning which is followed by a discussion of the use of computers to support language learning. I then discuss the current state of mobile assisted language learning. From a learning perspective, each of these three areas is organized around the roles that the learner’s exposure to language (hereafter referred to as language input), language production, and affect play. The additional role of adaptation and adaptivity within language learning, computer assisted language learning, and mobile assisted language learning is discussed because practitioners from all three fields have focused on how learning materials and experiences should be changed to meet individual learner needs.

2.1 Language Learning

While general educational theories can be employed to aid language learning, there are many theories that apply specifically to language acquisition and learning. These theories are based on previous research into various aspects of language learning and comprehension for both first language and other language learners. The methods that are often used tend to focus on theoretical frameworks that describe the specific types of activity that promote language acquisition.

Unless otherwise specified, the below section discusses language learning as an inclusive whole since there is considerable overlap between first language (mother tongue), second language, and foreign language learning. The type of language learning is specified when there is insufficient evidence to support its generalization across language learning groups or contexts.

Even though no single method seems to outperform the other, several frameworks and strategies that can be employed in language learning have been proposed. Learning strategies have been defined as any collection of activities that the learner can use to obtain, store, retrieve, or otherwise use information (Hardan, 2013; Oxford & Nyikos, 1989). This definition has been refined for language learning as those strategies that contribute to the language system that the learner has constructed.
These learning strategies can be supported through the use of drill and practice, problem-based learning, monitoring, or collaboration. Drill and practice obviously aims to support the rehearsal and memorization strategies of learners, whereas problem-based learning can be used to encourage the clarification or verification, guessing or inductive inferencing, or deductive reasoning strategies (Rubin, 1987). Both drill and practice and problem-based learning can be used to scaffold direct learning strategies by supporting the creation of mental linkages through the use of images, sounds, or actions (Oxford, 1990). Collaborative learning activities can support the communication, planning, or social strategies of learners; and monitoring activities, such as double checking one’s work for accuracy, help to support metacognitive strategies such as planning, prioritising, goal setting, and self-management. The collaborative and monitoring related strategies are sometimes referred to as indirect language-learning strategies and include the learner’s regulation of his or her affective state (Oxford, 1990).

Regardless of the language-learning or other learning strategies that are used, the learner’s exposure to and use of the target language influences his or her learning. The remainder of the discussion of language learning focuses on the roles that language exposure and production play within language learning. The below discussion also highlights the role of learner affect and individualization within language learning.

2.1.1 The Role of Language Exposure or Input

One’s exposure to language influences one’s knowledge of that language and one’s ability to learn that language, with the best predictor of mother tongue vocabulary acquisition being the quantity of speech heard and the frequency of word encounters (Wagner, Muse, & Tannenbaum, 2007). However, the language that one is exposed to must be meaningful for it to benefit learning. The noticing hypothesis further argues that language input only supports language learning when the person receiving the input is consciously aware of the language, i.e., notices it (Robinson, Mackey, Gass, & Schmidt, 2012; Schmidt, 1990). Noticing can be encouraged by teachers or through the use of journals (Kukulska-Hulme & Bull, 2009), and the language that serves as input can come from learning materials that are directly studied or the learner’s context.
2.1.1.1 Direct Study

Studies have revealed that language acquisition and the later production of language appears to be constrained by the person’s environment and that direct vocabulary instruction can assist acquisition (Wagner et al., 2007). If we consider vocabulary learning, it has been shown that all instructional methods provide more benefit than no instruction and that no individual form of instruction outperforms the others (Wagner et al., 2007).

Direct study can be guided as in the sage on the stage model, where an expert has organized material and delivers that information to learners. The information can be delivered through any communication mode (e.g., graphical, textual, or audio), but it rarely involves deep levels of interaction. Another commonly used method involves drill and practice and is relatively easy to implement. It is sometimes used after students have made mistakes on dictation tests. In these cases, teachers have students copy the word or sentence out a number of times.

In addition to this type of drill and practice, spaced repetition is also used to ensure that vocabulary items are learned (Pavlik & Anderson, 2008). This helps to ensure that words are learned by exposing learners to them at the appropriate times so that the information is stored in the learner’s long-term memory.

Regardless of which language-learning strategies are used, observations of learners have indicated that those "who are better at detecting and manipulating syllables, rhymes, or phonemes are quicker to learn to read" (Wagner et al., 2007). One technique that follows from this is the recommendation that vocabulary instruction include teaching external (surrounding text) and internal (within the word, e.g., suffixes, roots, and prefixes) context cues (Wagner et al., 2007). Furthermore, both children and adults are better at reading phonologically stable words (e.g., cultural) than they are at reading shift words (those that undergo a phonological transformation, such as natural) (Carlisle, Stone, & Katz, 2001). It is, therefore, important to use materials that combine audio and textual information because it can benefit learners by showing the relationship between the text and its pronunciation (i.e., it can help to create a mental linkage).

Combining the instruction of internal and external context with drill and practice has also been recommended (Wagner et al., 2007). This practice should include prompting students to rehearse
these skills while attempting to determine the meanings of words (Wagner et al., 2007). Learners should be taught about the multiple meanings of words while being encouraged to practice recognizing when different meanings are used (Wagner et al., 2007).

The above methods can help build metalinguistic awareness skills, which includes phonemic and morphological awareness. These types of metalinguistic awareness can account for a large portion of the variance in the correlation between test results for vocabulary knowledge and reading comprehension (Nagy, 2007). The development of metalinguistic awareness has also been shown to help bilingual children comprehend text (Gathercole, 2007). However, teaching metalinguistic skills presents risks. For example, students can confound the meaning of an unknown word with that of the entire sentence (Nagy, 2007) which means that this teaching must be done with care.

2.1.1.2 Learning from Context

Incidental learning is a form of learning that is unplanned and happens as the result of unintentional exposure to materials and activities (Jones, 1989; Kerka, 2000). It can happen in any educational environment and does not result from the direct instruction of the learned term. However, incidental learning shares characteristics with the methods that are considered most effective in formal educational settings (Rogers, 1997). That is, incidental learning is often social in nature, contextual, and situated.

Incidental learning is tied to the acquisition hypothesis which argues that how a person acquires knowledge is as important as the knowledge acquired (Schank, 1995). In language learning, it is often achieved through the reading of texts that are meaningful to the learner (Elley, 1997), with readers being able learn the meaning of words from one incidental learning opportunity (i.e., fast mapping) (Wagner et al., 2007). This has led to the suggestion that incidental learning is the primary method by which vocabulary is acquired until at least the end of primary school with it continuing into the later stages of life (Wagner et al., 2007).

Like incidental learning, situated learning is strongly tied to context (Lave & Wenger, 2011). It can be unintentional and states that learning depends on the context of the activity being performed. This theory argues that learning is useful when the method used to acquire knowledge is set in an appropriate social and physical context. In other words, teaching and
learning must occur in an authentic situation. This makes the use of ubiquitous technologies well suited to situated-learning practices.

Further to this, increased exposure to a wide variety of words in recurring contexts that are semantically rich appears to help with acquisition for monolingual and bilingual children (Wagner et al., 2007): encountering words in different contexts demonstrates the breadth of their meaning. For example, the word *port* could be a drink or the place where ships dock. The increased exposure to language use in different settings, including those where different registers (e.g., academic or legal English) are used has been shown to enable the language learning of recent adult immigrants (Gordon, 2004). Moreover, exposure to oral language is a foundational part of vocabulary acquisition, and it is the medium through which children first use fast mapping, a form of incidental learning, to understand the meaning of words (Carey, 2004, 2010; Wagner et al., 2007). This is also the method by which adults acquire vocabulary in the absence of instruction (Wagner et al., 2007).

In addition to exploiting higher-level contextual information, learners can use their knowledge of morphology to infer the meaning of words (Wagner et al., 2007). This can give learners a general idea of what a word means. However, learners may lack the knowledge necessary to interpret these clues, thus hindering their ability to use contextual information and infer the meaning of a word. In the sentence ‘He was playing with his fly’, the fly could be an insect or the zipper on the protagonist’s pants, additional information is needed to determine the meaning of the word *fly*.

**2.1.2 The Role of Language Production**

Existing theories of language learning claim that language production plays three roles: it allows us to notice gaps in our knowledge when we cannot communicate our intended message, it allows us to test our hypotheses about the language, and it enables us to reflect on our language knowledge and use (Swain, 1995). Studies that have been conducted since this theory was first proposed have provided evidence for language production supporting these three functions (Robinson et al., 2012). That said, language production, like language input, may only enable learning when the person is attending to their use of language and the responses that she or he receives (Robinson et al., 2012).
The production of oral language contributes to all forms of learning by helping us to organize, extend, and shape our thoughts (Jade, 2009). Since all communication, is to an extent collaborative, systems that support socialization or the user’s interactions with those in his or her surroundings can help enable learning by allowing the user to participate in language use and meaning creation (Gordon, 2004). This type of effortful language production is often called languaging (Swain, 2006). It has been shown to result in learner hypothesis testing and the reformulation of the learner’s understanding of the target language, which is consistent with the manner in which Levine (1966) and Anderson described learning (1993). In addition to changing a learner’s understanding of how the language is used, the use of interaction and language production can help to increase the authenticity of learning by situating the use of language in real-world contexts that are meaningful to the learner.

Other theories, such as social constructivism (Vygotsky, 1978) or languaging (Swain, 2006), also focus on how interactions between learners influence their learning (Paavola, Lipponen, & Hakkarainen, 2002). These theories emphasize the importance of having language learners interact with one another using the target language. Some of these theories further explain the influence that the surrounding environment has on learner behaviour (Bandura, 1986; Pajares, 2002): they state that the responses of those in one’s surroundings will influence a person’s willingness to perform a task or exhibit behaviours in a specific situation. This means that learners may be more likely to use a mobile application to study during their morning commute than they would if the other people on the subway were not playing games or checking email on their phones. People also model those with whom they identify (Bandura, 1986; Pajares, 2002). Therefore, the use of mobile tools by language learners is likely to encourage other language learners to use them. Beyond this, there is theoretical support for people’s ability to learn through observation and participation (Bandura, 1986; Pajares, 2002), which includes their ability to produce language to interact with others.

2.1.3 The Role of Affect

While affect can be used to refer to a person’s mood or emotions, I use it to describe both mood and emotions. I then specify whether I mean a temporary affective state (momentary affect) or a longer-term, trait-based affective state (Ekman, 1994). Because this work focuses on conducting research in the field, I further rely on the individual’s perception of his or her emotional
experience rather than a physiological measurement of a person’s affective state, even though both are possible.

I employ a commonly used definition of affect that consists of two dimensions (Picard, 2000; Russell, 1980): valence (i.e., the amount of pleasure being experienced) and arousal or activation (i.e., the amount of energy that is being experienced). If we plot affective states using a coordinate plane, the labels that we apply to individual states will fall within one of the four quadrants of the graph (Figure 4).

While it is recognized that affect influences learning and learning activities (Ashby, Isen, & Turken, 1999; Bixler & D’Mello, 2014; Bosch et al., 2015; D’Mello, Lehman, Pekrun, & Graesser, 2014; Hawkins, Heffernan, & Baker, 2013; Picard, 2000; Wen, Yang, & Penstein-Rosé, Carolyn, 2014), it has been operationalized and measured in many ways. The measurement approaches include sentiment analysis (Wen et al., 2014), observer coding (Baker, Corbett, Roll, & Koedinger, 2008; Bosch et al., 2015), self reporting (Clark & Watson, 1999; D’Mello & Graesser, 2012; Hektner, Schmidt, & Csikszentmihalyi, 2007), physiological measures collected through sensors (Heraz & Frasson, 2011; Lisetti & Nasoz, 2002), and multi-model approaches.
Regardless of the approach that is used, models have been built to explain or predict the relationship between affect and learning.

The above studies focused on the broader field of adaptive computer assisted learning. Recent studies into the role of affect in language learning provide evidence and additional theory regarding its role. Outside of learner anxiety, the role of affect within language learning has not been deeply studied. The most common view of the role between the learner’s affective state and learning is the affective filter (Swain, 2013). This theory argues for the relationship between language learning and the learner’s affective state (Dixon et al., 2012; Krashen, 1982; Swain, 2013) and at a high level states that a learner must be open to learning in order for learning to occur. Since the affective filter was originally proposed in 1977, people have been working on operationalizing the incredibly broad view of affect that it uses (Laine, 1987), with several different operationalizations of affect being shown to have a relationship with language learning (Krashen, 1982). Specifically, learner self-confidence and motivation support language learning and anxiety hinders learning (Dixon et al., 2012; Krashen, 1982; Swain, 2013). However, a more nuanced view of affect’s mediating role within learning may help improve our understanding of language learning and acquisition (Swain, 2013).

2.1.4 Learner Modeling and Adaptation

In formal educational contexts, instructors may build mental abstractions or models of a learner’s abilities and adjust learning materials or the support that is provided based on those models. Learners can also monitor their own learning activities and outcomes. They can then reflect over their activities and adjust what they are doing to achieve their goals.

One of the ways in which learning is adapted to a particular learner or group of learners is through the use of scaffolding. Scaffolding is the act of supporting a student’s understanding so that the learner can attain his or her full potential (Puntambekar & Hübscher, 2002). It is the support that is provided to students in Vygotsky’s zone of proximal development (ZPD), where the ZPD is defined as the discrepancy between what a learner can do on his or her own and what that same learner can do when given appropriate support (Vygotsky, 1978). Outside of formal
educational contexts, learners may have someone or something that can scaffold their understanding and activities, but this may not be true in all settings.

In language learning, this support can be achieved by reducing the difficulty of words or the length of passages so that they are appropriate for the learner (Wagner et al., 2007). This adaptive scaffolding can be even more effective when additional supports are used to reduce cognitive load by increasing the ease with which information access tasks can be performed (Sweller, Ayres, & Kalyuga, 2011). For example, errors can be reduced by providing the student with a list of word spellings at his or her desk instead of having the student look to one that is at the front of the classroom (Wagner et al., 2007).

2.2 Computer Assisted Language Learning

Computer assisted language learning (CALL) is the use of computational technology to support language learning (Beatty, 2010; Stockwell, 2012). It dates back to the 1950s and can be done through desktop or laptop computers and the programs that run on these devices. Some people include mobile devices, such as smartphones (Beatty, 2010; Garrett, 2009), in this category but I will treat them separately given that they afford different activities (Barkhuus & Polichar, 2010; Biancalana et al., 2011; Church & Smyth, 2008; Fogg & Eckles, 2007; Gay, 2009; Stockwell, 2012). CALL can be achieved through the adaption and adoption of existing technologies or through the use of dedicated programs and technologies (Beatty, 2010; Levy & Stockwell, 2006). CALL has been used across educational contexts: it has included learner use of these tools both outside and within classroom settings (Levy & Stockwell, 2006).

Many of the more widely used CALL tools that predate this century relied on fill in the blank or multiple choice type activities, and basic drill and practice approaches are still widely used (Garrett, 2009). We have only seen an increased prominence of more complicated, multimedia, approaches since the turn of the century (Beatty, 2010). However, the majority of these multimedia activities are still focused on exposing the learner to language, while increasing material authenticity and supporting learner comprehension with glossing and annotation tools (Beatty, 2010; Garrett, 2009; Stockwell, 2012).
Outside of the use of computer mediated communication (Beatty, 2010; Garrett, 2009; Levy & Stockwell, 2006), CALL tools typically fail to encourage the learner’s use of the target language. When systems encourage languaging and other production activities, learners rarely receive in-depth feedback unless a human-in-the-loop approach is used (Beatty, 2010). This may be the result of technical challenges that are related to diagnosing misconceptions and errors or providing appropriate feedback. However, there are some CALL systems that provide tutoring with limited foci (Demmans Epp & McCalla, 2011; Levy & Stockwell, 2006; Mitrovic, 2007; Price, McCalla, & Bunt, 1999).

Many desktop and laptop CALL systems have been incorporated into formal educational settings at various levels (Barrow, Mitrovic, Ohlsson, & Grimley, 2008; Beck, Jia, & Mostow, 2003; Demmans Epp & McCalla, 2011). CALL programs have been used in primary and secondary school settings and are heavily used at the post-secondary level with little attention being given to their efficacy (Garrett, 2009). That said, one study showed that students who used CALL in class learned more than those who did not (Chapelle, 2010). The use of CALL at the post-secondary level and for supporting less-commonly taught languages (e.g., Swahili or Cree) (Garrett, 2009) also means that a vast collection of CALL materials have been created (Levy & Stockwell, 2006) even if these materials have not always been evaluated (Chapelle, 2010).

The availability of CALL tools does not mean that they can be easily adopted or integrated into language-learning programs or practices. Arguments have been made and evidence provided for the need to train both teachers and learners in their use (Beatty, 2010; Garrett, 2009; Lai, 2015; Stockwell, 2012). Arguments have also been made for an increased focus on CALL pedagogy and design (Levy, Hubbard, Stockwell, & Colpaert, 2015).

2.2.1 The Role of Language Exposure or Input

Many CALL systems provide the learner with language input by exposing him or her to language through the use of drill and practice exercises or by glossing what are deemed to be authentic target-language materials. These are high-level examples of the use of CALL to support direct study or learning from context. However, few systems rely on only one of these approaches. Like in a classroom setting, it is far more common for some combination of approaches to be used.
2.2.1.1 Direct Study

Direct study was common in early CALL systems and can be seen in games like crossword puzzles or Rooting Out Words (*Rooting Out Words*, n.d.), where players must identify the correct word based on a description of a target word and the description of a morpheme (a meaningful word part, e.g., the suffix *-s* or the prefix *un*). Some direct study systems have also used spaced repetition to support vocabulary learning (*Teach2000*, 2011), and several text-based systems use drill and practice to support vocabulary and grammar learning (Martin & Nicholas, 2007; *Super Munchers*, n.d., *Teach2000*, 2011). A classic example of this type of system is Word Munchers, which was used in many North American primary schools during the 1980s and 1990s to help children improve their phonological awareness and increase their vocabulary (*Super Munchers*, n.d.). The development and use of these types of CALL is common partly because this medium and approach is similar to more traditional language learning practices for grammar and spelling. Moreover, it can be easily implemented.

2.2.1.2 Learning from Context

Some systems, especially those that are commercially available, use a variety of methods. Rosetta Stone is one that employs context, as conveyed through images, to support language learning ("Rosetta Stone," n.d.). Other, experimental systems have employed the content of the user’s chat to determine the user’s context and support his or her learning by displaying the target language label for words that are associated with the user’s current communication topic (Cai, Guo, Glass, & Miller, 2014). En una palabra tries to help its users acquire listening skills, vocabulary, and cultural knowledge of an area in Argentina by employing audio monologues and transcriptions of those monologues with additional supports, such as glosses (Pardo Ballester, 2009). Some of the more unique approaches to supporting language learning through context are the use of a simon-says like activity through a gaming console (Edge, Cheng, & Whitney, 2013) and the use of a simulation environment (Johnson, Marsella, & Vilhjálmsson, 2004).

2.2.2 The Role of Language Production

Many of the more recently developed CALL tools have supported learner production of language in a controlled way. Systems that use this approach to production limit the user’s opportunities for learning through language, but they enable the system to provide the learner with feedback.
There are exceptions to this when humans are not included in the feedback-generation loop. These exceptions will be covered in greater detail in the section on learner modeling and adaptation.

Systems that constrain user language production do so with prompts that require the learner to produce written or oral language that matches the prompt in some way (Demmans Epp & McCalla, 2011; Mostow, 1998; Papadopoulou, 2008; “Rosetta Stone,” n.d.; von Ahn, 2013; Wang & Seneff, 2007). This input could be the result of translating the prompt into the target language or another language with which the learner is familiar (Demmans Epp & McCalla, 2011; von Ahn, 2013), saying the prompt (Demmans Epp & McCalla, 2011; Mostow, 1998, 1998; “Rosetta Stone,” n.d.; von Ahn, 2013) or having the learner type what she or he has heard (von Ahn, 2013).

In some cases, CALL allows for more complex input (Johnson et al., 2004; Mitrovic, 2007). However, this is rare within applications that are dedicated to supporting language learning. More complex forms of input are typically used when technologies have been repurposed to support language learning (e.g., computer mediated communication tools such as blogs). The ability of different forms of computer mediated communication to consistently support learning gains (Stockwell, 2012) is likely due to the languaging activities that are supported through the learner’s interactions with others. Learners also gain production experience and feedback through their interactions with others across interactive tabletops (Paluka & Collins, 2013) or in online gaming platforms, such as Second Life and World of Warcraft (Garrett, 2009).

2.2.3 The Role of Affect

Much of the work surrounding affect in adaptive learning environments, independent of their instructional domain, has focused on its detection and modeling so that it can then be used to adjust instructional materials (Baker et al., 2008; Bosch et al., 2015; Hawkins et al., 2013; Lehman, D’Mello, & Graesser, 2012). Very little of this work has been performed in the area of language learning. Rather, within CALL, the role of affect has centered on supporting the learner’s affective state (Barrow et al., 2008; Demmans Epp & McCalla, 2011; Lai, 2015). One study showed that supporting learner affect had the greatest influence on learner uptake of CALL technologies for self-directed learning (Lai, 2015), while another provided evidence that the use
of appropriate feedback could help maintain learner motivation (Demmans Epp & McCalla, 2011).

2.2.4 Learner Modeling and Adaptation

Many CALL systems adapt learning materials and activities to learners with the aim of providing a personalized learning experience. These types of experiences are encouraged from a dynamic assessment perspective and rely on Vygotsky’s theory regarding the zone of proximal development (Poehner & Lantolf, 2005). Dynamic assessment focuses on the learner’s potential, and its use allows each student to receive the personalized support that he or she needs in order to reach his or her potential. Independent of the domain of learning or platform on which personalized learning is delivered, dynamic assessment requires that the system have knowledge of the learner that includes the learner’s background, beliefs, knowledge, intentions, goals, or abilities. This system-level knowledge is usually referred to as a student or learner model (Bull, Brna, & Pain, 1995; Bull & Kay, 2007, 2010). It can be thought of as the knowledge that a good teacher has about his or her student, and that the teacher can use to properly scaffold that student’s development. Learner models enable e-learning systems to adjust materials and activities, much like a teacher might, so that the learner can benefit from system use.

The type of adaptation that is encouraged by dynamic assessment is not always achieved. However, personalized learning experiences can be achieved through the use of spaced repetition algorithms (Elmes & Fraser, 2012), item-response theory, or a learner model (Barrow et al., 2008; Beck et al., 2003; Bull & Pain, 1995; Demmans Epp & McCalla, 2011; Price et al., 1999; Shahrour & Bull, 2008; Tao, 2007). In order for a system to develop a model of the learner, it must closely monitor his or her actions. This allows the system to assess or infer the learner’s knowledge, beliefs, or abilities. This tracking and the inferences that are made from the collected data then form the system’s learner model. In many cases, this involves the comparison of learner input against the expected correct input (Garrett, 2009; Kodaganallur, Weitz, & Rosenthal, 2005). When user input is more complex, its correctness may be determined by applying a set of constraints or other established modeling techniques (Demmans Epp & McCalla, 2011; Garcia, 2013; Martin, 1999).
The ease of comparing a user’s input to a correct answer, may be why many of the early and simpler adaptive systems for supporting basic language knowledge and skills employed a drill and practice or learning by testing approach (Bull & Pain, 1995; Jianguo & Xiaozhen, 1994; Mich, Betta, & Giuliani, 2004; Mostow, 1998; Nagamani, Narendra Prasad, & Girija, 2005; Tao, 2007; von Ahn, 2013). More complex systems have employed simulation-based approaches that provide learners with feedback about their appropriate use of language (Johnson et al., 2004; Price et al., 1999), but these are rare.

2.3 Mobile Assisted Language Learning

Mobile assisted language learning (MALL) has been defined in many ways. This includes any form of language-learning that uses mobile technologies, any form of language-learning where the learner is mobile, any form of language learning where the activity is mobile, or a combination of these (Hockly, 2013; Palalas, 2011). Regardless of the definition of MALL, some claim that the use of mobile tools by language learners is supplanting more traditional approaches to language learning, including the classroom (Beatty, 2013). Others claim that the use of MALL tools is still a fringe activity (Burston, 2014b) with its primary use being the review of previously studied content (Burston, 2014a).

While I would agree that MALL has the potential to supplant current language-learning approaches, considerable effort into improving the language-learning activities that are currently supported by these tools and the pedagogy surrounding their use would be necessary. Claims that MALL is in the early-adopter stage of the diffusion of innovation model (Y. Rogers, Sharp, & Preece, 2011) seem reasonable. I would additionally specify that MALL lacks the cultural norms and pedagogical acceptance necessary to move into the early majority or later stages of this model, which is supported by recent research (Burston, 2014a; Kukulska-Hulme, 2013; Sweeney, 2013). Educators appear to use MALL because of insufficient class time, varied learner abilities, excessively large class sizes, the need to differentiate their training programs, and the desire to use the equipment that students already have access to in the hope that this will enable learning activities to be continued outside the classroom (Sweeney, 2013).

Some proponents of mobile learning have argued that mobile computing allows learners to work towards personal learning goals and overcome challenges in addition to increasing learner
autonomy and providing them with anytime-anywhere access to learning materials (de Jong, Specht, & Koper, 2008; Kadyte, 2004; Liu, 2009; Ogata & Yano, 2004). There is evidence that learners will take advantage of this anytime-anywhere learning (Swan, van´t Hooft, Kratcoski, & Unger, 2005; Wong & Looi, 2010). However, social conventions may prevent anytime-anywhere access. Even with the addition of this potential constraint on anywhere-anytime learning, the use of mobile tools can allow for additional practice and learning opportunities. These opportunities can empower learners to use their time more effectively and achieve more (Edge, Searle, Chiu, Zhao, & Landay, 2011). For example, learners could use their daily commute to review materials but may only practice listening to pronunciation examples or speaking in the safety of their homes.

If we consider why language teachers said they promote the use of MALL (Sweeney, 2013), it is perhaps surprising that improving student learning experiences and outcomes was not among these reasons. The uses that they recommended also aimed to support receptive language learning tasks that were teacher led (Sweeney, 2013), which demonstrates a lack of integration of MALL into their curricula. If we consider MALL tool use in classroom settings, only electronic dictionaries and translators seem to be widely used. Other MALL use is still rare (Burston, 2014b). Excluding vocabulary support, the most common course-based uses of MALL have involved the repurposing of existing tools such as podcasts, audio recording, or mobile blogs to support language-learner activities (AIOkaily, accepted; Demouy & Kukulska-Hulme, 2010; Godwin-Jones, 2011).

While the field of MALL is established and there are many reports of explorations into mobile device and application use in formal and informal learning environments, several researchers have argued that teachers need guidance on how to select appropriate MALL tools (Duman, Orhon, & Gedik, 2014; Kukulska-Hulme, Norris, & Donohue, 2015; Son, accepted). There is a general lack of knowledge about how to best support learners with these tools and how to integrate MALL into the different types of classroom settings in which they can be used. Most MALL work has been performed in monolingual educational contexts where the target language is being studied as a foreign language (Hockly, 2013). This ignores many of the contexts in which language teachers work, including the North American context where classes contain people from varied cultural and linguistic backgrounds. However, appropriate guides that
synthesize prior research are beginning to appear in the form of example lesson plans and templates for creating lesson plans (Kukulska-Hulme et al., 2015).

The integration of mobile systems has been less explored and the literature provides less guidance than it does for CALL systems (Poole et al., 2011). Even though teachers lack sufficient guidance in how to best use these systems, they are in a position where they need to guide their students in MALL tool use because many students lack the necessary knowledge to do this on their own (Kukulska-Hulme, 2013; Lai, 2015). In order for schools to adopt these technologies, a strong case must be made for their potential benefit to learners, and this potential benefit must outweigh the potential distraction and disruption that the introduction of mobile devices can cause (Poole et al., 2011). Making a case for the benefit of MALL in classroom settings is problematic because the focus of many MALL as drill and practice or content delivery tools (Burston, 2014a; Hockly, 2013; Kadyte, 2004; Son, accepted) is disconnected from the current pedagogical preference for the communicative approach to language learning (Beatty, 2013). In addition, MALL tend to focus on a subset of theories from language learning (Godwin-Jones, 2010) and are still mimicking previous paper-based and computer-based approaches (Beatty, 2013; Kukulska-Hulme, 2013; Kukulska-Hulme, Sharples, Milrad, Arnedillo-Sánchez, & Vavoula, 2010; Stockwell & Hubbard, 2013). This is in spite of the opportunities, such as supporting collaboration or noticing, that current mobile technologies and the situations in which they can be used afford (Ballance, 2013; Burston, 2014a, 2014b; de Jong, Specht, & Koper, 2010; Kukulska-Hulme, 2013; Kukulska-Hulme et al., 2010; Stockwell & Hubbard, 2013; Vladoiu & Constantinescu, 2011; Wong & Looi, 2010).

MALL tools and their use have not been thoroughly evaluated beyond the explorations of their initial design and development (Burston, 2014b; Duman et al., 2014; Hockly, 2013; Wu et al., 2012). Many of these evaluations have taken place in artificial settings (Stockwell, 2010) and were relatively short (i.e., from a few hours to a few weeks) (Burston, 2014b). These evaluations have tended to rely on survey or observational methods with little tracking of learner activities or progress beyond the use of post-testing (Burston, 2014a; Wu et al., 2012).

Most MALL tools have only been evaluated in a single context or with lower-proficiency learners in post-secondary settings (Burston, 2014b; Wu et al., 2012). Few studies have included
MALL use by advanced learners, high school students, or immigrants who find themselves having to use the target language to survive (Burston, 2014b; Kukulska-Hulme, 2013; Traxler, 2013). Most studies of MALL and adaptive learning environments have been conducted within a single cultural setting and fail to account for differences in the cultural expectations of users (Blanchard, 2014; Burston, 2014b; Kukulska-Hulme, 2013). These limitations in the evaluation of MALL tools increase the difficulty of addressing teacher and administrator concerns about the effectiveness of MALL, which may be negatively affecting the integration of MALL into classroom settings.

2.3.1 The Role of Language Exposure or Input

The transmission model of learning still dominates MALL unless a teacher provides the learner with feedback on the artefacts that she or he has submitted through the MALL tool (Burston, 2014a). Several content delivery-based MALL have helped support the language learning activities of users and in some cases their learning (Dearman & Truong, 2012; Edge et al., 2011). Many of these systems have been based on exposing the learner to language in context, where context is defined using the device’s physical location, while others have allowed the learner to determine what she or he would study. The seeming popularity of building MALL that employ repetition to expose users to learning content and to ensure that vocabulary items have been learned (Alelo: Immersive simulations of real-life social communication, n.d., “FullRecall - Software For Effective Memorization,” n.d., “Welcome to the Mnemosyne Project,” n.d.; Dearman & Truong, 2012; Edge et al., 2011; Elmes & Fraser, 2012; Johnson et al., 2004; von Ahn, 2013) may be partially explained by the distinguishing feature of successful vocabulary instruction being the frequency with which learners encounter words (Wagner et al., 2007).

2.3.1.1 Direct Study

Many of the direct study tools resemble dictionaries (CloudBank, n.d.; Demmans Epp, Tzourounis, Djordjevic, & Baecker, 2013; Munteanu et al., 2010; Procter-Legg et al., 2012) or flash cards (Edge et al., 2011; Elmes & Fraser, 2012; von Ahn, 2013) and aim to support vocabulary learning (Duman et al., 2014). When systems resemble flashcards, they tend to use a spaced repetition approach to expose the learner to vocabulary. The systems where users can guide their own learning by selecting the categories that they will study also tend to keep
learning activities short (Edge et al., 2011; Elmes & Fraser, 2012; Procter-Legg et al., 2012). The MobiLearn electronic phrasebook provides an example of a system that allows learners to review words and phrases as well as assess their knowledge through tests (GrapeCity, 2001).

2.3.1.2 Learning from Context

Several systems rely on some form of context to support the learner’s ability to study (CloudBank, n.d.; Dearman & Truong, 2012; de Jong et al., 2010; Demouy & Kukulska-Hulme, 2010; Edge et al., 2011; Kadyte, 2004; Pincas, 2004; Son, accepted; Vladoiu & Constantinescu, 2011). Many of these systems claim to be context-aware but are actually location-aware and present learning materials based on the user’s GPS information. Others provide context through the use of stories. This form of context is most common when podcasts are being used.

Situating language learning in the learner’s everyday context and bringing that learning back into the classroom was performed in a Chinese primary school, where students documented real-world occurrences of the Chinese idioms that were being studied in class (Wong & Looi, 2010). This activity included a follow-up group exercise where learners discussed their results and reflected on how well the documented observations of each idiom matched its intended meaning. Other researchers and practitioners have implemented similar programs that require learners to notice and log interesting usages of language in their environment (Demouy & Kukulska-Hulme, 2010; Kukulska-Hulme & Shield, 2008; The Raft Project, 2004). This noticing and logging of language has employed augmented reality approaches to label real-world artefacts (Santos et al., 2014). It has also included the use of contextual information, such as culture in the case of LingoBee and its predecessor Cloudbank (Procter-Legg et al., 2012).

2.3.2 The Role of Language Production

Few of the MALL that are dedicated to language learning support production. When teachers repurpose mobile tools to support language production they tend to have learners record and submit a production task through the mobile device (Burston, 2014a). When MALL tools explicitly support production, it tends to be on limited tasks such as the pronunciation of a word or phrase (iSpeaky, 2011; Munteanu et al., 2010; Son, accepted; von Ahn, 2013) or the identification of words on a grid that contains characters (Jain, Birnholtz, Cutrell, & Balakrishnan, 2011). In some cases (e.g., LingoBee), users may contribute to a larger community
by adding additional information about regional language that they have noticed (CloudBank, n.d.; Procter-Legg et al., 2012) or use the system to share their experiences and obtain assistance from fellow learners (Ogata & Yano, 2004).

Few tools encourage interaction with other speakers. The MALL tools that encourage interaction in the target language use collaborative learning approaches to support primarily text-based interactions (CloudBank, n.d.; Ogata & Yano, 2004; Procter-Legg et al., 2012) rather than oral communication. That said, there is a system that aims to support communicative language learning outside the classroom by enabling the rehearsal and submission of speech (Liu, 2009).

Some applications that could be considered MALL are support tools whose purpose is to minimize the burden placed on the user’s memory system. Support tools such as PixTalk (De Leo, Gonzales, Battagiri, & Leroy, 2010), MyVoice (“MyVoice,” 2011), or Proloquo2Go (“Proloquo2Go,” 2009) can support language learning activities even though they are intended to support communication (Wong & Looi, 2010). All three of these example applications display text alongside a visual representation of its meaning (De Leo et al., 2010; “MyVoice,” 2011, “Proloquo2Go,” 2009), which can affect vocabulary acquisition by creating memory linkages that assign the target-language label to a concept or object that is known to the learner.

2.3.3 The Role of Affect

Very little work in the area of MALL has focused on the relationship between system usage and user affect. When affect is considered, it is usually only considered from the perspective of learner preparedness or acceptance of mobile technologies (Hsu, 2013; Stockwell, 2010; Stockwell & Liu, 2015). On the rare occasion that investigators are interested in determining the influence that a MALL tool has on learner affect, they rely on qualitative surveys and interviews (Wu et al., 2012). Whereas, the influence that communication support tools have on user affect is often reported (Kagan & Simmons-Mackie, 2007) but does not allow for easy comparison between studies because it is qualitative in nature.

2.3.4 Learner Modeling and Adaptation

Mobile devices are especially well suited to enabling personalized learning because they are viewed as personal devices (Barkhuus & Polichar, 2010) and their owners usually keep them
nearby. This section focuses on the adaptation that happens within MALL independent of a teacher’s intervention. There has been very little overlap between the exploration of the use of intelligent-tutoring-like or other personalized approaches and MALL (Hung & Zhang, 2011) even though these are occasionally used within other mobile-learning domains. Adaptivity within MALL tends to follow a location, item-response-theory, or spaced-repetition model (Burston, 2014a; Dearman & Truong, 2012; Edge et al., 2011; Elmes & Fraser, 2012; Stockwell & Liu, 2015). However, the use of individual learner models and other approaches to adaptation and personalization, while rare (Traxler, 2013), is becoming increasingly common.

For example, Duolingo offers users the option of performing a placement test that is used to refine the learning content from which the user can choose (Garcia, 2013; Vesselinov & Grego, 2012). Duolingo is also one of the few systems that supports all four language macro-skills (i.e., reading, writing, speaking, and listening) while using sentence level context to increase learner exposure to words. This design appears to have helped it support beginner learners (Vesselinov & Grego, 2012), but Duolingo’s effectiveness for supporting more advanced learners is still being questioned as a result of a deterioration in the quality of the system’s feedback as the language becomes more complex and the proficiency of learners more advanced (Garcia, 2013). Its effectiveness is also questioned as a result of demonstrations of the superiority of the translations provided by Google Translate (Garcia, 2013).

In rare cases, situational and environmental cues (e.g., available time and likelihood of being interrupted) are included as part of the context-modeling process (Cui & Bull, 2005). This information is then employed to adjust learning materials to the user’s situation so that she or he is given learning tasks that are appropriate to the learning environment and the user’s language knowledge. More commonly, the user’s GPS location is used to adjust materials: one system used internet-based corpora to identify noun-verb pairs that were relevant to the user’s current location and show the translations for these word pairs to users (Dearman & Truong, 2012). Project InLET also exploited the user’s location to support learning; it enriched the learning experience by delivering content through several media forms (MMS and SMS, leaflets, broadcast programs, and posters) during the Greek Olympic Games (Introducing Language Enhancement Techniques: InLET, 2003)
There are only three adaptive MALL that can also be used as support tools (VocabNomad, Proloquo2Go, and the Handheld Companion). All three of these applications employ user models to adapt to the user. VocabNomad provides adaptive learning content and vocabulary (Demmans Epp et al., 2013), and the Handheld Companion adapts its learning activities to the user (Alelo: Immersive simulations of real-life social communication, n.d.). Proloquo2Go adapts the items presented to the user based on those recently used, and it adapts the user’s input to ensure that compositions are grammatically correct (“Proloquo2Go,” 2009). Of these three systems, VocabNomad is the only one that provides both communication support and adaptive learning materials.

2.4 Conclusion

Some MALL projects have supported the learning activities of students through the repurposing of existing applications that are intended to support other activities. Specifically, the use of computer-mediated-communication applications, such as mobile blogs, have been shown to support language learning. Other MALL tools have provided the learner with personalized support and learning activities that could be completed at the learner’s convenience. Of these dedicated applications, a few have been shown to support short-term gains in the learner’s vocabulary knowledge.

More commonly, the studies of these MALL tools have been performed from the perspective of evaluating their design or feasibility. Prior work shows little evaluation of MALL tools for supporting learner affect beyond his or her preferences. The evaluation of the influence of MALL tools on learning outcomes is rare even though several surveys have shown that computer-based intelligent tutoring systems for language learning and other domains have undergone considerable evaluation for their influence on learning outcomes (Ma, 2008; Stockwell, 2012; VanLehn, 2011). We have also seen little work tying the use of MALL, user affect, and user learning together.

Like other educational technologies (Blanchard, 2014), MALL systems are not typically evaluated across contexts. However, the use of MALL has been studied in a wider array of cultural and educational contexts than many other educational technologies even if a single system is only evaluated in one context.
Few existing MALL scaffold oral communication and only a handful of MALL scaffold written
communication. The focus on knowledge transmission or grammar-translation approaches within
MALL also means that limited forms of adaptivity have been employed within MALL tools and
that they have not been evaluated for their ability to support communication.

These limitations informed the approach that was taken to evaluating the MALL technology that
we built. These limitations also informed and led to the three primary research questions that are
addressed within this thesis.

**RQ1.** To what extent can an adaptive mobile tool support both the vocabulary
learning and communication of ELLs?

**RQ2.** How do users integrate an adaptive mobile tool into their language
learning and communication practices?

**RQ3.** What is the relationship between the usage of an adaptive mobile tool and
ELL vocabulary knowledge, communication success, and affect?

The answers to these questions will be explored following a discussion of the development of a
MALL tool that uses a new approach to support ELLs. This approach to scaffolding language-
learning is based on that provided by communication support tools. As a result, we begin with an
exploration of the use of a communication support tool by ELLs (Chapter 3).
Chapter 3 – ELL Use of a Mobile Communication Support Tool

Since there have been few reports of English language learner use of communication support tools in spite of their potential usefulness for ELLs, we were interested in seeing how an adaptive mobile tool could support the vocabulary learning and communication of ELLs (RQ1). We, therefore, conducted a proof-of-concept study where we gave ELLs access to a communication support tool called MyVoice\(^1\). MyVoice had been designed to support the communication of those who had difficulties communicating for themselves as a result of a medical condition called aphasia (Demmans Epp et al., 2011). I set out to determine if ELLs could use it to support their everyday needs because the tool had been designed to support the communication of people during their everyday activities rather than in classroom settings.\(^2\)

This initial formative evaluation also aimed to determine which ELL communication and learning needs were not being met by their current support tools or the language learning systems that they had used. This study was the first step in the development and validation of the adaptive mobile assisted language learning system that is the focus of this thesis.

This proof-of-concept study focuses on evaluating the system as whole and was used to inform the design of a new system. This evaluation is exploratory in nature since the use of these types of mobile communication support tools had not been explored even though they hold the potential to support the learning activities of ELLs (Demmans Epp, 2013). The MyVoice tool was chosen because it was adaptive and considerable work had been performed with respect to refining its visual design in a way that enabled MyVoice to support the communication of those who had difficulty processing symbolic information in the form of language (Demmans Epp et al., 2011). MyVoice’s use had also been explored for its ability to provide cognitive support within special education contexts where students struggled with language comprehension (Campigotto, McEwen, & Demmans Epp, 2013; Demmans Epp, McEwen, Campigotto, &

\(^1\) Formerly MarcoPolo, now TalkRocket Go (http://myvoiceaac.com/).

\(^2\) This work appeared in Demmans Epp and Baecker (2012), Demmans Epp (2013), and Demmans Epp (2014). See Appendix L for a complete list of publications and their relationship to my thesis work.
Moffatt, 2015), which provided some evidence for its potential to support language learning. However, this evidence was insufficient so the below study was conducted.

3.1 Instrument: MyVoice

MyVoice is similar to the visual dictionaries that ELLs sometimes use, but it can adapt its support based on the user’s location. It is a dual-platform tool that has a web-based interface and an accompanying smartphone application (Figure 5). Like a dictionary, MyVoice does not directly support collaboration through either of its interfaces; it is intended to be used by an individual. However, others, such as teachers, could add vocabulary to a learner’s account if that learner decides to share his or her account credentials with others.

We deployed the initial version that ran on iOS devices; it had been iteratively evaluated and refined. Its user-centered design and development involved the continued use of the application by an adult male who was living with aphasia (a language disorder that disrupts one’s ability to communicate with others) (Demmans Epp et al., 2011). The application’s development included the solicitation of feedback from those who work with people who face communicative challenges as a result of developmental or medical conditions. Discount evaluation methods, including heuristic evaluation (Nielsen, 1994) and the application of the Gestalt principles of visual design (Mullet & Sano, 1995; Ware, 2004) were also used (Campigotto et al., 2013).

The mobile application interface delivers support materials and allows users to navigate through those materials. It does not allow the user to add new content and the only content that it allows the user to modify is the image that is associated with a vocabulary entry. The materials are organized based on their association with a physical place or their location within a hierarchy. The application development team created the initial hierarchy and users could modify the

![Figure 5 MyVoice's web interface (left) and mobile interface (right).](image)
hierarchy through the web-based interface.

From a MALL perspective, MyVoice was a talking visual dictionary, where each vocabulary entry was accompanied by an image representing its meaning. MyVoice allowed users to communicate by showing vocabulary entries to someone or by having the application verbalize those entries. MyVoice could also use GPS location information to provide vocabulary that the user had associated with a place. The web-based interface was used to enter vocabulary items, organize those items, and associate them with locations. While a regular web-browser was the only technology that was needed to add or edit content, the design of the interface and its reliance on drag and drop functions would have made these tasks difficult on a mobile phone.

3.2 Methodology

Participants were recruited through posters in public places and a government-funded language-training program that is offered through the Toronto public school system. The ethics approval letter for this study can be seen in Appendix A.

Participants came to the laboratory to complete a demographics questionnaire (Appendix B), see MyVoice’s functionality demonstrated, obtain the application on an iOS device (an iPod Touch 4 or iPhone 3GS), and receive their log on credentials. Participants were also given a user manual for the application and they were instructed to take the iOS device and application home to use in whatever manner they felt was beneficial.

After at least three weeks had passed, participants returned the device and were asked to participate in an interview. Interviews lasted approximately one hour, were recorded, and later transcribed. A shared paper-based writing space was used to facilitate communication (Figure 6). The example of the shared writing space that is shown in Figure 6 demonstrates some of the ways in which this space was used. The bottom-center of the page shows where the interviewer drew two possible interpretations of what the participant was saying. The participant then circled the sketch that best matched what he was meaning. Similarly, the line drawing at the top was used to clarify how he interpreted the difficulty of using different technologies and language learning tools. As we can see, he found it difficult to understand English television programs when the subtitles were in English, found his English classes less hard than the television
programs, and thought that using a computer was easy. I also had those who spoke a language that shared cognates with the languages that I know write what they wanted to communicate. I would then use this information to guess the intent of their message and confirm what they meant. The example from Figure 6, shows a phrase where the participant was trying to tell me that he could not or did not know how to explain something. With some participants, this shared writing space was used to sketch objects or situations. Other strategies, such as the use of dictionaries, translators, and charades were also used to overcome the language barrier that, at times, existed between the interviewer and participants. All participants were interviewed by the same person, who had experience as a second language (L2) educator; the language used was adjusted to each person’s abilities.

I conducted the interviews which were semi-structured and focused on five areas of personal experience:
• L2 learning: the strategies they have employed, the challenges they have faced, the successes they have experienced, and the aspects of language learning that they think are easy or difficult or that they have liked or disliked.
• L2 communication: the strategies they use to communicate with those around them, examples of successful and failed communication, how they have managed miscommunication, and examples of communication failures or successes.
• Technology use: the technologies they have used, how they have used them, and what they liked or disliked about these technologies.
• CALL and MALL technology use: the language learning technologies they have used, how they used them, and what they liked or disliked about them.
• MyVoice Use: when and how they used the application, what they liked or disliked about it.

The interview guide consisted of several higher-level questions and prompts that could be used to encourage conversation about one of the above-listed foci (see Appendix C). Participants were free to guide the conversation, provided they discussed topics that related to their language learning or integration within a culture where the dominant language of communication was not their mother tongue.

The original interviewer transcribed the interviews. The transcripts were then reviewed to identify the strategies that participants used, the technologies that they had used, and how they used MyVoice. Quotes that illustrated findings from previous research on language learning, depicted the strategies that participants used to learn, described the participant’s technology use, or disclosed the success or failure of communication with others were identified. Cards containing excerpts were created by including some of the surrounding discussion to maintain the context in which the quoted utterance was made (Singleton & Straits, 2010). These extracts were further anonymized and given to two other researchers for coding and thematic analysis (Creswell, 2013).

These coders sat at the same table and worked together to identify themes and grouped the excerpts accordingly. The original interviewer then reviewed the groupings and themes with the coders. At which time, the identified themes were discussed and refined to ensure that consensus
was achieved and that all excerpts and codes were interpreted consistently by everyone. The unmet communication and language learning needs of participants were extrapolated from the participant reports that fit within each of the identified themes.

3.3 Participants

Twelve ELLs from various linguistic and cultural backgrounds participated. The mean age of participants was 43.8 years (SD = 12.9). Their English language abilities ranged from functional fluency to only having the ability to answer simple questions, such as ‘What is your name?’ or ‘Where are you from?’ Details about participant demographics can be seen in Table 2. Participants’ current English language proficiency level on the Canadian Language Benchmark (CLB) scale was also reported (“Benchmarks (CLB),” 2006); this was only possible for participants who were enrolled in government training programs. CLB Level 1 describes someone who cannot decode or understand unfamiliar words; someone who is at level 4 in all areas (speaking, reading, listening, and writing) could be described as functionally fluent.

<table>
<thead>
<tr>
<th>ELL</th>
<th>Age</th>
<th>Sex</th>
<th>Mother Tongue</th>
<th>English Proficiency**</th>
<th>Language(s) Spoken at Home</th>
<th>Language [proficiency**]</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>49</td>
<td>M</td>
<td>Chinese (C.)</td>
<td>CLB 1</td>
<td>English</td>
<td>Spanish [little] &amp; Korean [little]</td>
</tr>
<tr>
<td>GS</td>
<td>42</td>
<td>M</td>
<td>Farsi</td>
<td>40% (CLB 4)</td>
<td>Farsi</td>
<td>Japanese [good]</td>
</tr>
<tr>
<td>M*</td>
<td>47</td>
<td>F</td>
<td>Spanish</td>
<td>Good (CLB 2)</td>
<td>Spanish &amp; English</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>18</td>
<td>F</td>
<td>Chinese (C.)</td>
<td>Fluent</td>
<td>Chinese (C.)</td>
<td>French [elementary]</td>
</tr>
<tr>
<td>O</td>
<td>44</td>
<td>M</td>
<td>Spanish</td>
<td>Poor (CLB 2)</td>
<td>Spanish</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>65</td>
<td>M</td>
<td>Bulgarian</td>
<td>Bad (CLB 3)</td>
<td>Bulgarian &amp; Russian</td>
<td>Russian [fluent]</td>
</tr>
<tr>
<td>P</td>
<td>36</td>
<td>F</td>
<td>Spanish</td>
<td>Poor (CLB 2)</td>
<td>Spanish</td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>46</td>
<td>F</td>
<td>Chinese (M.)</td>
<td>Good (CLB 4)</td>
<td>Chinese (M.)</td>
<td></td>
</tr>
<tr>
<td>Q*</td>
<td>55</td>
<td>M</td>
<td>Spanish</td>
<td>Poor (CLB 2)</td>
<td>Spanish</td>
<td></td>
</tr>
<tr>
<td>QS</td>
<td>55</td>
<td>F</td>
<td>Chinese (M.)</td>
<td>Poor (CLB 4)</td>
<td>Chinese (M.)</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>48</td>
<td>M</td>
<td>Chinese (M.)</td>
<td>Good</td>
<td>Chinese (M.)</td>
<td>Chinese (C.) [good]</td>
</tr>
<tr>
<td>S</td>
<td>21</td>
<td>F</td>
<td>Chinese (M.)</td>
<td>Good</td>
<td>Chinese (M.) &amp; English</td>
<td></td>
</tr>
</tbody>
</table>

* Spouses
** Self-reported, CLB levels were verified
+ Participants with 2 character identifiers, ending in S, used the application in the spring. All other participants used the application in late fall.
M: Mandarin, C: Cantonese

3 These participants were enrolled in courses that were part of the Language Instruction for Newcomers to Canada (LINC) program
http://www.tdsb.on.ca/AdultLearners/LearnEnglish/LanguageInstructionforNewcomertoCanada.aspx
All participants received 10 dollars in exchange for their participation even if they did not complete the final interview. Participant GS declined remuneration.

3.3.1 Participant Backgrounds

Participant GS had previously emigrated from Iran to Japan, where he learned Japanese by interacting with those around him which he enjoys and finds useful. He had no previous training in Japanese prior to immigration but had received some English training before moving to Canada. He is taking English classes and working in the kitchen of a Persian restaurant because it was the only job that he could obtain without better English proficiency. GS has learned and incorporated many of the pragmatics of Canadian English language use. He is functionally fluent and strategically competent: he can compensate for his lack of ability in other areas. For example, he described barberries, a sour berry that is used to season food, when he did not know the English word.

Participant N was no longer trying to improve her English abilities. She had learned English in her primary and secondary schools in Hong Kong and has achieved oral fluency. She still needs some vocabulary support and is pursuing an undergraduate degree at an English language institution.

Like N, S is pursuing an undergraduate degree at an English language institution and began her English training in primary school. She demonstrated functional fluency and strategic competence. She is trying to increase her fluency and understanding of academic vocabulary by using text-based tools and collaborating with classmates.

Participant R began learning English before immigrating to Canada in 2003. Since moving, he has taken university-level courses in English and struggled with academic language. He is no longer trying to improve his L2 abilities, but he had wanted to learn by talking with L1 speakers even though he struggles to understand them.

While PS is currently taking English courses, she appears to approach her learning in a relaxed manner and relies on her daughter to perform much of her L2 communication. She hurt her leg during the study and said that this was why she took a vacation. As a result, some of her data is incomplete.
Participants Q and M did not complete the full interview. Q and M are married and have two daughters who help them with their English. Their youngest daughter, who is not yet school-aged, takes pride in her ability to help her parents learn English. For more serious matters, they ask their nearly adult daughter to help them. Q and M handled L2 communication as a team. When something was said to them, they would discuss the message (in Spanish) and then one of them would respond. They each seemed to have different areas of strength and responsibility. For example, M handled all communication involving numbers.

In contrast to Q and M who worked as a team, OS allows his wife to perform almost all of his English communication. He is most interested in receiving support and shows little interest in learning to communicate in English. He seemed to have accepted that he will not achieve a high level of English proficiency, as is demonstrated through his claim that his inability to remember words may be due to his having sclerosis.

Participant O has a child who is fluent in English. They had lived in the United States for five years, prior to immigrating to Canada, where he had worked as a middle manager even though he struggles with communicating in English. He was surprised that he could not find work in Canada and was motivated to learn. He demonstrated strategic competence and inadvertently relied on the use of cognates to bolster his vocabulary during communication (e.g., explicarte for explain – see Figure 6).

Another Spanish speaker, P, was also unaware that she relied on cognates to communicate with English speakers. She is strategically competent and was volunteering for a non-profit organization where she could interact with L1 speakers of English. She had previously worked as a programmer but could not find work because of her language abilities and a lack of recognition for her credentials and previous experience.

Participant G had learned no English prior to immigrating. He is highly motivated and willing to try anything to learn. He is the only participant who did not rely on technology to support his efforts. He did not own a television or computer but was comfortable using them. He reported that he occasionally used his friend’s computer to access email.
QS used technology at home. Her ability to understand English far exceeded her ability to produce it, which is consistent with her reliance on passive learning strategies: she does not try to understand how the language works or investigate learning materials on a deeper level. She is often content with being able to understand English on a surface level.

3.3.2 Participant Language Learning and Language Use Summary

Six of the 10 participants who completed interviews reported that they had difficulty holding conversations and speaking, and 3 of the 4 remaining participants reported that grammar and writing were difficult: “Write is hard? Ya, I don’t like grammar”⁴ (GS). Regardless of the mode of language production, all participants thought that producing language was more difficult than understanding it. They also reported that a lack of vocabulary knowledge was one of the major hurdles that they faced when trying to understand language and when trying to produce language.

3.4 Exceptional Events

Four participants demonstrated behaviours that stood out from those of the group as a whole.

Participant GS felt that it was important to draw the interviewer’s attention to the words that he thought should be added to MyVoice because he used them at work every day. It was the first thing that he did upon arriving at the interview and he gave us permission to use the images that he had provided to improve the vocabulary that comes with the application. These actions are consistent with the desire for socio-collaborative learning opportunities, but the sharing of materials between users was not supported by the provided application.

One participant, QS, was concerned about her ability to communicate what she thought about MyVoice. As a result, she dictated her thoughts to her son before coming to the interview and provided a letter describing both how she had used MyVoice and what she liked or disliked

⁴ All quotes are word for word transcriptions from participant interviews.

*Int.* indicates interviewer utterances

/ / denote a description of performed actions

/ / indicate modifications to preserve the enclosing sentence’s grammar

/ / indicate phrases that the participant emphasized
about it. We reviewed this together and she was asked to expand on some of the items. She said that she had done this because her “English no good. Is because lot of new words for [her]. [She] didn’t know how to say”. This demonstrates that some learners, especially those at lower proficiency levels, need the opportunity to plan and even rehearse communication before they attempt it.

Participants N and P were highly aware of how they approached language use. Both knew when they were translating messages in their heads either to understand them or to communicate information to someone else. As N said, “you’re actually internally translating every word”. A similar conversation led to P revealing an awareness about the meta-cognitive aspects of her language learning and progress: “I like when I dream in English because I understand my brain is doing something, a step important.” This demonstrates that some learners may benefit from tools that help support their meta-cognitive abilities or allow them to track their progress.

3.5 Results

3.5.1 Learning Strategy Use

All participants had previously taken English courses and all but 2 (N and S) were currently taking English language courses. However, N and S were enrolled in degree programs at institutions where English is the language of instruction.

In addition to pursuing instruction, all participants used learning strategies that would develop their receptive language abilities. Everyone reported that listening to authentic examples of English language use was helpful. In one case (N), this involved listening to the conversations of others in the learner’s environment. More commonly, it involved listening to the radio, watching movies, or watching television.

All of the participants who mentioned using media to rehearse their listening skills used a combination of media. The most mentioned forms of media included television (with or without subtitles), radio broadcasts, and movies. Participants G and QS reported using dictation tasks to test their listening skills. N, P, and R joined conversation clubs that involved interacting with other L2 and L1 English speakers. Participation in these clubs was used to rehearse or test their listening skills and to develop their speaking skills. N, P, and R were also the only participants
who reported experiences that demonstrated an explicit attempt to develop their productive language abilities.

Participants N, O, P, and S relied on reading to improve their English and said that they enjoyed reading in English as long as the materials were at an appropriate level. They would read children’s books, internet articles, and news articles. Beyond the use of reading to learn English, S scaffolded her English comprehension within science classes by reading her textbook before attending lectures. N, O, P, and S also used a dictionary to enhance their understanding of English.

3.5.2 Previous Technology Use

Overall, participants were comfortable with technology and had used various technologies for language learning purposes. Participants reported the regular use of television and computers as well as the regular viewing of movies. Participant R said that the “computer like indispensable. It’s a requirement for life” when responding to questions about his technology use. Participants also listened to the radio or music on a regular basis and most participants (see Table 3) currently used mobile devices with at least two of them using smartphones.

While participants had used technology to support their learning activities, few of these technologies were dedicated to enabling language learning and none of the previously used technologies provided the user with additional instruction or scaffolding (see Table 3). Rather, participants had appropriated commonplace technologies to support particular aspects of language learning. They did this by changing how they used those technologies so that the technology met their needs (Dourish, 2003). For example, participants would use sub-titles to check their understanding of dialog in a movie or television show. While this example is a relatively common practice for language learners it demonstrates how technology use can be expanded beyond its originally intended purpose.

The language-learning technologies that participants had used enabled very little interaction. Participants QS, R, and S had used language-learning tapes but none of the participants had used any form of instructional software on either mobile or computer platforms. None had ever used adaptive learning technologies or communication support tools prior to the study.
### Table 3 Participant Technology Use – MyVoice Proof-of-Concept Study.

<table>
<thead>
<tr>
<th>ELL</th>
<th>TV</th>
<th>Radio</th>
<th>Music</th>
<th>Movies</th>
<th>Computer</th>
<th>Mobile Phone</th>
<th>Tapes / CD</th>
<th>CALL / MALL</th>
<th>Electronic Dictionaries</th>
<th>Paper Dictionaries</th>
<th>Electronic Translators</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>GS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>N</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>O</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>OS</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>P</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>QS</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>S</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Some of those who had used electronic dictionaries also used paper dictionaries; Participant G was the only one who used a paper dictionary exclusively. Participants also reported having used electronic translators on mobile devices. The electronic translators that I witnessed being used were used in the same way as paper L1-L2 dictionaries are used: participants translated individual words.

### 3.5.3 Participant Needs and Desires

Seven recurring themes emerged from the interview data. Most fall under three categories: 1) social and affective aspects of L2 learning, communication, and acculturation; 2) higher-level aspects of language usage within communication; and 3) learning approaches and required knowledge.

#### 3.5.3.1 Social and Affective Aspects of L2 Learning, Communication, and Acculturation

Participants focused on social and emotional experiences during their interviews. This revealed their need for social and emotional support, their frustration and discouragement with L2 learning, and their desire for socio-collaborative learning opportunities.
**Desire for social and emotional support:** Participants expressed gratitude towards L1 speakers who demonstrated patience and understanding during interactions with them. They were especially grateful when others seemed to understand their particular situation.

O: The people immigration. It’s good people. Ya, they understand the problems. My problems.

Participants also expressed frustration over the lack of patience and uncooperativeness of other people, especially service industry workers: “No, only teacher helpful” (G).

**Discouragement about L2 learning and communication:** Participants expressed frustration over the difficulty associated with learning certain components of English, and participant O was unable to identify any aspect of learning English that was easy. He simply said, “It’s all hard”.

Some participants were especially discouraged by the difficulties posed by their everyday lives. Participant OS relied on his wife to communicate in English on his behalf. Another participant, QS, expressed frustration over her inability to obtain L2 proficiency when her son and husband appeared to use English with ease.

**Desire for socio-collaborative learning opportunities:** Many participants wanted more opportunities to interact with L1 speakers. It was felt that this would help improve their ability to understand oral language and their ability to speak. In the absence of L1 speakers, participants wanted to practice their language use with classmates who could help clarify concepts or develop their skills.

N: Some people are better at the writing part but not so good at the speaking part, which is what happened to one of my friends, and I was better at the speaking part than her. So, I was, like, helping her, like, speak more fluently, and she helped me with the writing part.

R demonstrated a desire for socio-collaborative learning opportunities by expressing his belief that having closer relationships with more L1 speakers would have helped him learn non-academic English because it “is different setting. Right. Like normal, like, people you talk to in working place, school, or study place more formal subject. They don’t want to share and then if you get closer you have a lot of lively words or lively expressions”. This also demonstrates a desire for acquiring cultural knowledge so that it can later be used during communication.
3.5.3.2 Higher-level Aspects of Language Usage within Communication

For the most part, participants were not concerned with the lower-level aspects of communication (e.g., grammar and word order). Instead, they were focused on their ability to communicate as a whole and their need for the planning and practice that could improve their competence.

**Desire for L2 communication support:** When discussing communication support, it became clear that the ability to obtain support was more important in certain settings than it was in others.

GS: Go to bank, go to restaurant, go to /hospital/.

Int.: Okay. Hospital, specifically?

GS: Doctor okay. Everyone here has a doctor who’s same country, but hospital no.

Int.: Hospital’s more urgent?

GS: Go to hospital, maybe nobody does. You /have/ to speak English.

This example highlights both sides of communication: production and comprehension. It also provides a setting in which obtaining support for production and comprehension is especially important.

**L2 production support:** participants described the many tools and strategies that they used to support their ability to produce language (e.g., start a conversation or answer a question) and communicate with others. They also expressed frustration over the mismatch between their ability to understand and respond to those with whom they have interacted. QS said “I hear, maybe understand, the word. I speak. I don’t know how to speaking”.

**L2 comprehension support:** Participants wanted additional support when consuming language so that they could understand what is being communicated in audio or visual formats. Many participants communicated that they have difficulties understanding speech, especially when “people speak too fast and then [they] get lost” (S).

Participants were better able to understand static media, such as textbooks. However, participants expressed different media preferences that could limit their abilities to understand the language to which they are exposed. Some preferred using audio materials to support their understanding,
while others preferred to use text-based materials (e.g., S wanted lecture transcripts because “sometimes I know the words but when someone speaks them I can’t recognize that”) or visual supports (e.g., QS showed the application and stated, “Very good picture. Study there”). In some cases, participants also communicated distaste for audio media because it is more difficult to understand than other media.

Regardless of their media preferences, all learners needed some support to fully understand the language that they had encountered. Participant N only needed support when encountering new vocabulary items and struggled with determining the meaning of words based on their context of use. Unlike the other participants, she could understand the core of a message and later use a dictionary or the internet to figure out any of the message parts that had been unclear. Other participants needed higher levels of comprehension support.

Participants showed gratitude when describing situations where appropriate support had been provided. Participant O appreciated when one of his son’s teachers could use some Spanish during meetings because this let him understand his son’s academic progress, which had not always been possible following their immigration to Canada.

*Desire for rehearsal opportunities:* “I guess practice. Practice is always the best for everything” (N).

Beyond production support, participants wanted additional opportunities to rehearse their language use by listening, writing, or speaking. Some were comfortable with repeating the phrases that they heard on language learning tapes or that were presented through the application, whereas others wanted to discuss personal or professional topics with L1 speakers. In one case, a participant was frustrated that he could not rehearse his language use by interacting with those in the service industry.

GS: In Canada, you call any company and you can’t speak English. Always can wait. We can speak your language. It’s very bad, this country.

Int.: Because it lets you speak your language?

GS: For example, I go to the bank. I can’t speak English or I speak not very well. ‘Okay. Which language you speak?’ Persian. Somebody comes or helps. It’s fine but not good for me.
Those who did not feel ready to interact with L1 speakers wanted additional rehearsal opportunities to enable their later interactions.

S: I think it would be better if not only one sentence but conversations.
Int.: Is that because you like to model conversations?
S: Ya, I think that on that program what I should say but it would be better to add what someone response to that.

This desire to rehearse and prepare is also evidenced by how participant QS prepared for the interview that she knew would occur at the end of the study: she brought a typed letter that had been written in English.

3.5.3.3 Required Knowledge and Approaches to Language Learning

All participants recognized that they needed to learn certain aspects of the target language. They also acknowledged that incidental learning could be beneficial but that they had trouble gaining the exposure to English that is necessary for promoting the incidental learning that is based on language use in context.

*Necessity of Learning Vocabulary:* Not surprisingly all participants expressed, either in words or through reports of their activities, that their ability to use language was dependent on their vocabulary knowledge and that limited vocabulary knowledge made English language understanding and use difficult.

Int.: What do you find easy?
GS: Some words, in my country, I used.
Int.: So, you used some English before you came. So, that part was easy?

Participants described the many strategies and supports (i.e., dictionaries, translators, and Wikipedia) that they employed to build their vocabulary and understand those around them. For example, two of the Spanish-speaking participants (O, P) used cognates to understand and guess the word that they needed, whereas the Farsi participant (GS) would provide examples until the person with whom he was speaking guessed the word. If this failed, he would use an electronic translator to clarify his intent.
Participants also described situations where their attempts at communication had failed because of their limited vocabulary. For example, S was unable to understand her dentist when she went for a checkup because “the words [were] too professional”.

**Required and incidental learning**: Participants saw the necessity of learning English given their new geographical and cultural environment. Different media and examples of English language use were sought by participants. This demonstrates their understanding that exposure to high-volumes of English-language material can benefit their learning.

Some thought that immersion was the best way to learn: N stated that she thought that “just speaking, being in a different country” was the most beneficial to her language learning. This shows a reliance on the incidental learning that is enabled through one’s environment and through interaction with others.

One participant (GS), in particular, was frustrated by his inability to obtain L2 language exposure because of the amount of L1 support that was provided in his environment. He felt that this support was hindering his English development because he was not being forced to use his L2. Participant OS expressed a similar frustration because too many of the other students in his L2 program would not speak in English: “Ya. Too much Russian. People Afghanistan but speaking Russian”; he faced similar challenges at work where the majority of his coworkers also spoke Russian.

3.5.4 Application Use

Following deployment, none of the participants reported problems with using the devices, but some application-specific limitations were reported (e.g., the inability to translate words, record samples of language use, or get feedback about one’s pronunciation). Even with these limitations and the application’s inability to meet all of the participants’ communication-support or language-learning needs, participants reported that it was useful. Participant S described the application as convenient and when asked to define convenient said, “There are some categories on it. There are very good to know. Restaurant something that you could see. Then if you go there you can ask someone for help”. Only participant N did not find that the application was useful. However, she had achieved native-like oral proficiency.
In addition to the ease with which participants could use the device, none of them reported having problems with the application’s user interface. Those who felt that they could benefit from its use adopted it. Participant GS told a friend about the application and stated, “I have to buy this one because it’s very good” and some participants wanted to keep the software: “I take it uh uh home” (G). Participants G, GS, and a friend of GS’s also volunteered to help test future versions of the application. This provides evidence towards a desire and potential need for these types of tools.

Participants used the application to review and study the meaning of words (Figure 7). They also used it for listening practice, supporting their communication, and practicing the pronunciation of words and phrases: “I use it. Uhhh. I just tap it and it could pronounce that sentence. So, I, uhh, learned all of the sentences” (S). Some participants would go a step beyond this to test their phonetic decoding skills and vocabulary knowledge by performing dictation tasks as is evidenced by a report from participant G: “I only look here {demonstrated his application use}. I listen. Listen. Here, I writing. I writing. Listen. Listen”.

Although participants liked the portability of the application because “you can use it wherever you are” (P), they primarily used it in the privacy of their homes and did not use its location aware features. Participants O, P, QS, and GS used the application in public spaces. For GS, this meant that he was silently reviewing vocabulary while riding the bus or subway when commuting between his home, work, and school. The others used it with friends or as a prompt to support their communication.

In one case, participant P, who is communicatively competent but speaks softly and has a strong accent, reported using the application to support her communication in a real-world situation where service-industry staff did not understand her and were being impatient. If she had not had the application she would have had two options for dealing with this situation: “you repeat and you motion and frustrate … other strategy is to go to the other kind of stores”.

51
Figure 7 Participant use of MyVoice to support their learning and communication goals.

Interestingly, those who were confident trying to communicate in their L2 felt that MyVoice could support vocabulary acquisition but not communication even when their ability to communicate was limited. This is not to say that they viewed the application as just an electronic picture dictionary even if they primarily treated it as such. In fact, participant S stated that the application “is not like a dictionary”.

It appears that participants viewed the application as a content delivery tool even though they did not believe it to be a dictionary. Participants did not seem to perceive the application as one over which they could exert complete agency. This is supported by the limited amount of learning content modification and creation that participants performed, with only GS identifying vocabulary items that should be added. However, we do not know if this perception is the result of the application’s design or participants’ experiences with other applications, the majority of which are likely to only allow interactions with predetermined content.
3.6 Discussion

Like the learners who used the HELLO system (Liu, 2009), our participants readily accepted the use of mobile devices for learning English. Learners used the MALL application that they were given to explore and acquire new knowledge when it was most convenient for them. Our participants’ desire for additional opportunities to rehearse their language use and communicate with others supports Liu’s claim that most learners thought that using English in real-life situations could improve their L2 knowledge and their ability to produce language (2009). When combined with our participants’ desire for socio-emotional support and previous findings that learners may prefer to practice interactive speaking exercises in private locations such as the home (Demouy & Kukulska-Hulme, 2010), this study confirms the need for a safe practice environment (Demmans Epp, 2010; Vosniadou, 2001). This can be seen in our participants’ repurposing the application to support speech and pronunciation rehearsal in the privacy of their homes where the social norms that may otherwise prevent rehearsal can be disregarded.

The occasional use of the application’s speaking function by some participants also confirms previous investigations that demonstrated its ability to support communication (Demmans Epp et al., 2011). However, the fact that the majority of participants did not attempt to directly support their communication in this way may indicate shortcomings in the vocabulary support that was provided by the application or the presence of social barriers that prevent the tool’s use for this purpose. It may also indicate that ELLs require training in how to use these types of tools to support their communication.

While MyVoice may not have been useful for the only participant whose oral proficiency resembled that of native English speakers, all other participants felt that it supported their vocabulary learning activities. This support even benefitted those whose proficiency was high enough to gain admission to academic institutions where English is the language of instruction (i.e., participant S).

However, participants wanted additional information and features. Some of the features that participants wanted were similar to those desired by participants in another study (de Jong et al., 2010). Participants from both studies shared a common desire to access samples of vocabulary use in context. However, our participants expressed a need for additional context beyond that of
individual sentences to support their rehearsal and later communication; De Jong et al.’s participants only wanted sample sentences to support their understanding of individual vocabulary items. Our participants did not express this explicit need, but they had access to phrases that showed how some words could be used in context. Furthermore, many of their experiences demonstrated that the use of sample sentences had benefitted their vocabulary acquisition and language comprehension in the past. It has also been shown that this type of exposure to vocabulary increases knowledge through the fast-mapping and extended mapping processes (Carey, 2010; Wagner et al., 2007) and could increase the usefulness of providing socio-collaborative learning opportunities, especially those that involve L1 speakers.

While few of our participants viewed the provided application as a tool that supported communication, it was used as a topic of discussion between friends and, on occasion, to support the task-based communication of learners who had lower L2 proficiency levels. This does not contradict previous findings that mobile devices can be used to support social conversation (de Jong et al., 2008), but indicates that they can also support task-based communication. When combined, these results indicate that MALL applications can be used to support socio-collaborative learning as well as the L2 learner’s ability to produce language.

Our participants’ desire and need for help in understanding language extends previous explorations that focused on knowledge acquisition (Dearman & Truong, 2012; de Jong et al., 2010; Edge et al., 2011; Liu, 2009) and the recording of language use in situ (Demouy & Kukulska-Hulme, 2010). Furthermore, participant preferences for using static media (i.e., visual and textual aids) and their occasional distaste for transient information (e.g., audio media) are consistent with the types of support that are suggested by cognitive load theory: unknown information should be presented in static media to best support its retention (Sweller et al., 2011).

That said, several participants used the application to develop their listening comprehension and provide a model of English pronunciation. The provided application is slightly better suited to this than the television or radio because it pairs audio with textual representations of the word or phrase, which could help learners develop their phonetic decoding skills and contribute to reading comprehension (Carlisle et al., 2001). The practice of listening to a pronunciation model
while reading the accompanying text could also improve vocabulary acquisition and comprehension if it succeeds in increasing the learner’s awareness of English graphotactic patterns (Deacon, Conrad, & Pacton, 2008). Additional increases in orthographic knowledge could occur, especially for those who perform dictation practice. Increased orthographic knowledge may result from dictation practice because it can increase productive phonological rule awareness (Carlisle, 1988) and later improve one’s ability to speak.

In contrast to De Jong et al.’s results (2010), our participants showed no interest in organizing the provided materials. This may be due to the learning content being well organized. However, the reasons behind this difference are unknown and require further exploration to explain why some learners want this ability and others refrain from adding to the learning content or modifying existing content when given the opportunity.

3.7 Unmet Participant Needs

While MyVoice supported basic communication (i.e., placing an order at a coffee shop), it was insufficient for supporting communication in all settings (RQ1). Participants expressed concerns over the lack of vocabulary to support communication in unforeseen circumstances or high-stakes settings, such as the emergency room at a hospital. This scaffolding is needed to support the social and affective needs of ELLs with respect to their ability to take advantage of socio-collaborative learning opportunities. It could also improve the support that they receive for the basic knowledge that is required in everyday life.

Beyond the ability to obtain support when they need it, participants expressed a need for recording the language that they encounter in their daily lives as well as the language usage of others. Participants wanted to receive feedback on the accuracy of their pronunciation, which fits with their desire to record the language usage of others to gain access to a pronunciation model to which they can aspire.

Based on these needs and the needs that can be derived from the above results and discussion, a set of recommendations can be made for the creation and improvement of MALL tools. These recommendations can be used in applications that are designed to support any of the above-
identified themes. However, their specific implementation may vary based on the goals or learning objectives of the particular MALL application.

These recommendations include

**DR 1.** Pair static media with transient media, such as audio or video, to support the development of more than one macro skill. This pairing of media will also support the learner’s ability to control the extraneous cognitive load that is imposed by the learning and support materials.

**DR 2.** Provide support for emergent learner needs.

**DR 3.** Allow learners to record novel language use and add to or modify existing learning materials.

**DR 4.** Provide opportunities for learners to plan and rehearse their language usage.

**DR 5.** Provide the learner with feedback.

**DR 6.** Allow learners to track their progress and reflect over their learning and communication. At a minimum, do not disable features that allow the learner to perform self-monitoring activities (e.g., the dictation practice that was performed by participant G).

**DR 7.** Enable the use of translation services where appropriate. Do not enforce a target-language only model of L2 acquisition. Learners should be able to use the other languages that they know to support their L2 learning activities.

**DR 8.** Allow for learning activities that can be performed in public spaces as well as those that can be performed in private spaces.

**DR 9.** Provide the learner with contextual and usage information where appropriate.
3.8 Conclusion

An evaluation of a location-aware cognitive and communication support tool by English language learners was performed to determine if ELLs could benefit from this type of support. The evaluation ($N = 12$) revealed that ELLs could use the tool to support specific language-learning activities (RQ1 and RQ2) but that there were several factors that impeded ELL use of the tool. These factors as well as the various learner needs that were not being met through other support systems were distilled into a set of nine design recommendations. These design recommendations were then used to develop a system (see Chapter 4 and Chapter 5) that implements a new approach to supporting ELLs with MALL tools: this approach scaffolds ELL communication, provides support for emergent user needs, and dynamically recommends learning materials to users. The development of the user interface and scaffolding mechanisms will be described next.
Chapter 4 – VocabNomad: User Interface Design and Adaptive Feature Validation

Following the proof-of-concept study that was conducted using a location-aware communication support application (Chapter 3), a new tool was designed to better meet ELL needs. The design of this new tool was informed by the theories that are detailed in Chapter 2 and the results of the proof of concept study. This tool is called VocabNomad and its design included the development of a feature to address the unmet need of ELLs for vocabulary support in emergent situations. This feature, just-in-time or on-demand vocabulary support, provides the user with new learning and support materials based on an identified need. The developed on-demand vocabulary support consists of two components (word list generation and image matching) that were validated in separate user studies. The design of the system and its user interface incorporated the design recommendations that had been identified as a result of the proof-of-concept study (see Chapter 3), while maintaining MyVoice’s basic visual design elements because ELLs seemed to appreciate many of the aspects of that system’s base design. The newly designed user interface also underwent validation.

This approach to system validation, hereafter referred to as the layered approach, is consistent with a framework that decomposes interactive adaptive systems into different conceptual layers (Paramythias, Weibelzahl, & Masthoff, 2010). Each layer corresponds to a stage in the adaptive process and is comprised of the elements that contribute to the system functionality for which that layer is responsible. This evaluation framework shows how the components from each of the stages or layers of adaptivity can interact to provide the adaptation that may or may not be visible to the end user (see Figure 8).

This framework includes the following layers:

- Collect input data: data is obtained through non-interactive sensors (e.g., keystroke logging, thermometers, or accelerometers)

- Interpret data: the raw data that was collected in the previous layer is given meaning within the context of the system.
Figure 8 The layers of the evaluation framework and their interactions.
(from Paramythi et al. (2010); used with permission – see Appendix D).

- Model the current state of the world: new knowledge about a user, interaction context, or other modeled component is derived. This is where knowledge of the world is shared with the dynamic model that is part of an adaptive system. The dynamic model is updated based upon the user’s interactions with the system or changes in the surrounding environment.

- Decide upon adaptation: this is where the system decides if adaptation is required based on the underlying models. This is also where the appropriate type of adaptation is chosen.

- Apply adaptation: based on the decisions made in the previous layer, adaptations are performed.

Depending on the manner in which a system is developed, each component of an adaptive system may belong to more than one layer. Regardless of the system’s structure, errors from one conceptual layer can propagate to subsequent layers; thus, magnifying errors in system
reasoning. These errors can reduce the quality of the user’s experience or, in the case of adaptive educational technology, harm learning.

The MyVoice proof-of-concept study from Chapter 3 was used to inform the design of the interactive “front end” layer. This chapter focuses on the discourse completion and Internet-based studies that were used to validate the just-in-time vocabulary support, which is part of the “static” domain model. This domain model directly contributes to the interpret data layer and indirectly contributes to the apply adaptation layer. Heuristic evaluation of the mobile tool’s user interface and café studies were used to partially validate the user interface and the adaptive theory component of the apply adaptation layer. They were also used to partially validate the interaction history portion of the “dynamic” models component. The dynamic models component is part of three layers: model the state of the world, decide upon adaptation, and apply adaptation. The final formative evaluation (Chapter 6) of VocabNomad was used to validate the system as a whole and to verify that the interaction history from the dynamic models component was receiving the correct information from the collect and interpret input data layers.

The design recommendations (Chapter 3) that resulted from the MyVoice proof-of-concept study inform the system’s design and provide some evidence for its validity. Its use by learners will be illustrated via personae in Chapter 5. Here, I describe the relationship of the design recommendations to the redesign process. I then discuss the additional steps that were taken to validate the user interface before presenting the development and validation of the on-demand vocabulary creation that was added in response to DR2 (provide support for emergent learner needs).

4.1 Initial (Re)design

To better support the situated learning for which mobile tools are ideally suited, the presentation of learning and support materials was modified. A sentence showing the use of the word or phrase within a larger context was added. This is meant to foster incidental learning processes that allow the learner to obtain a partial understanding of a word’s meaning the first time that it is encountered in context (i.e., fast mapping) and refine their understanding of that word during subsequent encounters (i.e., bootstrapping and extended mapping) (Carey, 2004, 2010). This practice meets the ninth design recommendation (DR 9). Moreover, it is linked to the learner’s
later ability to use the vocabulary item appropriately (Godwin-Jones, 2010) and is a feature of other MALL tools, such as LingoBee (Procter-Legg et al., 2012).

DR 1 continues to be met by the pairing of speech synthesis (i.e., text to speech) with the text that is provided by the application. The decision to allow learners to control when text to speech is used takes into account their need to limit the cognitive load that the support materials create as well as their ability to control which types of activities they want to perform based on where they are performing those activities (DR 8). The addition of a feature where learners can record audio samples of language usage also meets this need and builds on DR 3 and DR 1 to allow learners to record regional variances in pronunciation. This feature may resemble the audio note-taking feature of some other tools but is semantically different in that it is intended to capture learning content rather than a learner’s commentary about learning content, as is the case with LingoBee (Procter-Legg et al., 2012). The playback feature of the audio recording supports DR 4, 5, and 6 by allowing learners to record themselves saying something and then playing it back to compare their pronunciation to that of the expert, which is provided by the text-to-speech engine. The playback feature also allows ELLs to continue rehearsing a target phrase until they are happy with their performance.

It was decided that additional support for DR 5 and 6 would be added later. The design and development of a type of feedback mechanism called an open learner model has been started (Tsourounis & Demmans Epp, in press; Tsourounis, Demmans Epp, & Baecker, 2014), but the feature was not included as part of the version of VocabNomad that was evaluated because it had not been sufficiently validated when the first summative evaluation of the system as a whole was conducted. Moreover, including the open learner model would have meant that the final system evaluations were performed using different versions of VocabNomad. This would have limited the comparisons that could have been made between the two studies.

To address DR 3, we included the ability to add new vocabulary entries or edit existing entries through the mobile application interface, which supports learners’ noticing activities (Kukulska-Hulme & Bull, 2009; Robinson et al., 2012; Schmidt, 1990). We scaffold user’s emergent needs (DR 2) by allowing learners to obtain learning and support materials on demand. The addition of this just-in-time support is one of the distinguishing features of VocabNomad.
Table 4 A summary of the changes that were made to the basic system design as a result of the proof-of-concept study, its associated design recommendations (DR), and the extent to which the design recommendation is supported (ES).

<table>
<thead>
<tr>
<th>System Area</th>
<th>MyVoice</th>
<th>VocabNomad</th>
<th>DR (ES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content presentation</td>
<td>• image-word pair</td>
<td>• added a sentence</td>
<td>9 (Sml.)</td>
</tr>
<tr>
<td></td>
<td>• text-to-speech</td>
<td>• added ability to mute</td>
<td>1 (Lge.), 3 (Lge.), 4 (Sml.), 5 (Sml.), 6 (Sml.), &amp; 8 (Mod.)</td>
</tr>
<tr>
<td>Content organization</td>
<td>• hierarchical organization</td>
<td>• added ability to make and play recordings</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• association with locations</td>
<td>• added on device content editing and creation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• web editable only</td>
<td>• graph-based organization</td>
<td></td>
</tr>
<tr>
<td>Content access</td>
<td>• navigation through hierarchies</td>
<td>• added search</td>
<td>2 (Mod.) &amp; 7 (Sml.)</td>
</tr>
<tr>
<td></td>
<td>• location-based refinement</td>
<td>• navigation by category</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• added on-demand content-generation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• added translation</td>
<td></td>
</tr>
</tbody>
</table>

Sml. – Small, Mod. – Moderate, Lge. - Large

We also decided to change how vocabulary entries were organized. VocabNomad uses a graph-like organizational structure rather than a hierarchical organization. This was done to reduce data redundancy, enable on-device editing of learning materials, and support learner schemata development with respect to the various relationships between vocabulary items. It holds the potential to increase the learner’s breadth and depth of knowledge of particular vocabulary items by making the relationship between the word and various categories visible.

The above changes are summarized in Table 4, which previews some of the other modifications that are explained in greater detail later in the chapter. These modifications include the substitution of an explicit search feature for the location-based vocabularies that were intended to reduce the user’s search space. This decision was based on a lack of feature use in the formative evaluation and evidence from a study where students who used the location-based feature had repurposed locations to flatten the navigational hierarchy so that they would not have to search through the ontology to find the words that they wanted (Demmans Epp et al., 2015). The changes that are summarized in Table 4 required considerable modification of the user interface. I worked closely with a graphic designer to modify the visual layout of information, to ensure

---

5 The extent to which a design recommendation is supported is a descriptive judgment that I made. It is meant to indicate how well the design recommendation is supported but is not an absolute measure of the application’s support for that design recommendation.
consistency in the semantics of visual elements across platforms, and to employ icons that communicated system functionality. A discussion of the validation of the above-described user-interface changes follows.

4.2 User Interface Validation

Initial user interface validation involved the application of Neilson’s heuristics (Nielson, 1994) and the Gestalt principles of visual design (Ware, 2004). Interface elements that were inconsistent with these principles were identified and modified while keeping the context of use, user abilities, and the system’s objectives in mind. One example of a change that was made as a result of this process was the modification of the web-interface’s delete button: its colour was changed to a light grey from bright red to reduce its contrast with the rest of the system and prevent users from clicking on the button as a result of it being given greater visual importance than the surrounding buttons (see Table 5). Since the initial designs in Table 5 were only intended to represent the visual layout and allow for its refinement, random content of appropriate lengths was entered for each of the types of data that were to be included.

Café studies were performed to validate the user interface of the mobile application since they are a type of formative evaluation that is used as part of the professional practice of designing interactive systems on mobile devices (Konno & Fong, 2013). Café studies are very brief case studies in which the designer or developer enters a coffee shop, approaches the individuals who are in that coffee shop, and asks them to complete a highly constrained task. A sample task might include asking a participant to edit the sentence that is associated with a particular vocabulary entry or asking a participant to find a particular word within the system. You then watch how several people respond, and you adjust the design of user interface elements that are associated with tasks that participants had trouble completing.

The user interface was also adjusted to meet the recommendation that MALL systems accommodate learner differences and their existing cultures of use (Stockwell & Hubbard, 2013). This accommodation was provided through the addition of a search feature. Differences in learner abilities and needs were addressed through the use of different types of media: text to speech and recording for audio media, pictures as a form of visual media, and text. Learners were given control over which features were used at which times to accommodate their existing
cultures of mobile device use. This control allowed them to mute the verbalization or text-to-speech feature should the learner be in an environment where the use of audio is inappropriate (see Table 6).

Table 5 Some of the stages in the design and refinement of the web-based user interface.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td><img src="image1" alt="Initial Design" /> <img src="image2" alt="Initial Design" /></td>
</tr>
<tr>
<td>Intermediate</td>
<td><img src="image3" alt="Intermediate Design" /> <img src="image4" alt="Intermediate Design" /></td>
</tr>
<tr>
<td>Final</td>
<td><img src="image5" alt="Final Design" /> <img src="image6" alt="Final Design" /></td>
</tr>
</tbody>
</table>

These were the two initial design options that were created based on the sketches that I gave to the graphic designer. The design on the right is the one that was chosen as the base template.

Colloquial language was removed from the interface (e.g., Trashed was changed to Deleted). The colour, size, and placement of some buttons were also changed.

Paging was added and the images and functions were made more consistent with those of the platform and the rest of the system.
Table 6 Some of the stages in the design and refinement of the android client user interface.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>The original MyVoice communication support tool design is on the left. The designer-created mock-ups of the main screen (centre) and detailed screen (right) that were based on my sketches and our discussion are shown.</td>
</tr>
<tr>
<td>Intermediate</td>
<td>The recommendation of learning content was added (left). The ability to record oneself or others was added (right). The visual communication of the edit function was made consistent with the web platform.</td>
</tr>
<tr>
<td>Final</td>
<td>Access for the pronunciation models was moved to the detailed view so that all audio features and their control were in the same place. This also removed the overloading of the semantics for a touch on the main screen.</td>
</tr>
</tbody>
</table>
Table 5 and Table 6 show some of the designs that were considered and should enable readers to see some of the changes that were made as a result of the above described evaluation and validation. These tables do not show all of the adjustments that were made, but Table 5 provides an impression of how the user-interface changed over time. Similarly, Table 6 starts with the initial MyVoice design that served as the basis for VocabNomad’s user-interface and gives a sense of how the user-interface was continually refined.

4.3 On-Demand Vocabulary Support: Development and Validation Studies

To meet participants’ expressed need for vocabulary support in unplanned or emergent situations (DR 2), the ability to request new support materials was developed. Since this requires the creation of considerable quantities of content (Traxler, 2013) and we cannot expect our users to do all of this, an automated approach was selected for providing this just-in-time or on-demand support. That is, the system needed to be able to generate collections of appropriate learning materials. These support materials needed to provide vocabulary that could scaffold communication given the user’s current context. The meaning of the provided vocabulary also needed to be communicated if the materials were to be of use to ELLs. It was decided that images would be used to fulfill this need because they reduce the system’s reliance on natural language to communicate vocabulary meaning. The methods used to generate support materials (i.e., the algorithms that generate vocabulary collections for an identified context and the methods that find visual representations of a vocabulary item’s meaning) were evaluated separately. Their evaluations are described below.

4.3.1 Generating Collections of Vocabulary On-Demand

Information-retrieval algorithms were applied to Internet-based corpora with the intent of providing support for emergent learner needs (DR 2). The vocabulary collections that were generated by the information-retrieval algorithms were evaluated for their ability to support communication that was specific to a learner-identified context or need (Demmans Epp et al., 2012). This evaluation also served to validate the domain model that contributes to the interpret data and apply adaptation layers of the layered approach.
4.3.1.1 Instruments and Procedures

A discourse completion study (Nurani, 2009) was conducted but its administration was modified. Participants were asked to respond to situations as best they could by using the vocabulary that had been generated by four different algorithm-corpus pairs (Table 7). Participants selected words from a list to compose their written response to the stimulus. The algorithms that were evaluated are based on three basic information retrieval approaches: term frequency (tf) (Jurafsky, 2009), term-frequency-inverse-document frequency (tf-idf) (Salton & McGill, 1986), and the relative location of words algorithm (Carpena, Bernaola-Galván, Hackenberg, Coronado, & Oliver, 2009). Additional details about the algorithms can be seen in Appendix E and the training and study discourse tasks that participants completed can be seen in Appendix F.

Participants were allowed to use blanks if the provided vocabulary did not meet their needs. They then later completed their message by typing in the missing words to fill in the blanks. Examples of participant responses can be seen in Table 8. The discourse completion tasks required participants to make requests and negotiate misunderstandings in the contexts of shopping, ordering food and beverages in a restaurant, attending movies at a theatre, and working.

4.3.1.2 Participants

Seventeen participants were recruited through postings on bulletin boards and online forums; 16 completed the study. The one who withdrew had completed approximately 25 percent of the discourse completion tasks an hour into the study so we stopped his participation. All reporting provides information on the 16 who completed the study.

<table>
<thead>
<tr>
<th>Table 7 The algorithm-corpus pairs that were evaluated.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Dictionary</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Website</td>
</tr>
<tr>
<td>Wikipedia-USF</td>
</tr>
</tbody>
</table>
Table 8 Examples of discourse responses that could still be understood, even if words had not been provided (missing words) by the algorithm-generated vocabulary. The actual response is the message that participants could communicate using the algorithm-generated vocabulary lists, the intended message shows how the participant had wanted to communicate that message.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Situation</th>
<th>Intended Message</th>
<th>Missing Words</th>
<th>Actual Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Purchasing tickets</td>
<td>Hi. I'd like to buy one ticket.</td>
<td>to buy</td>
<td>Hi. I’d like one ticket.</td>
</tr>
<tr>
<td>3</td>
<td>Purchasing tickets</td>
<td>One ticket, please</td>
<td>please</td>
<td>One ticket</td>
</tr>
<tr>
<td>10</td>
<td>Ordering drinks</td>
<td>Glass of white wine please</td>
<td>Glass, of, please</td>
<td>white wine</td>
</tr>
<tr>
<td>9</td>
<td>Refusing help</td>
<td>No, thank you. I am just looking.</td>
<td>Just, looking</td>
<td>No, thank you.</td>
</tr>
</tbody>
</table>

Their average age was 33.9 years ($SD = 11.7$). All participants were fluent in English and were educated to at least the high school level in English. Participants received 10 dollars for their participation. One participant also requested and received a summary of the study results.

4.3.1.3 Results

To ensure that the discourse completion tasks were balanced, we first determined the response length for each of the tasks by counting the number of words that were in participant responses. Testing revealed that there was no significant effect of algorithm or theme on the response length (in number of words). This may indicate that the vocabularies generated by each of the algorithms had little effect on response length (see Table 9).

To make sure that the individual discourse tasks did not affect the length of the messages by requiring a more polite set of responses for a particular theme or location, we tested to see if there was a politeness effect in participant responses. We tested for politeness because it can

Table 9 Mean and standard deviation of response lengths in number of words.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Sites</td>
<td>5.70</td>
<td>2.81</td>
</tr>
<tr>
<td>Dictionary</td>
<td>5.73</td>
<td>2.65</td>
</tr>
<tr>
<td>Location's Website</td>
<td>5.83</td>
<td>2.77</td>
</tr>
<tr>
<td>Wikipedia-USF</td>
<td>5.97</td>
<td>3.10</td>
</tr>
</tbody>
</table>
increase word count by adding additional politeness markers, such as please and thank you. It can also increase the length of a response by decreasing directness, e.g., “I would like to pay for two tickets for Friends with Benefits” (participant 8) rather than “Can I have two tickets for Friends with Benefits”. A two-way ANOVA revealed that there was no significant effect of politeness (amount of indirect speech or use of politeness markers) on response lengths for each theme or algorithm ($p = .10$). This indicates that the types of responses required by the discourses were balanced.

As expected, there were a number of individual differences observed in participant responses. Chief among them was a significant difference between the length of individuals’ responses as shown by the results of a two-way ANOVA test ($F_{63,255} = 32.91$, $p < 0.01$, $\eta^2 = 0.92$) which also revealed that the politeness levels used by participants were significantly different ($F_{63,255} = 9.40$, $p < 0.01$, $\eta^2 = 0.75$) even though they did not vary across discourse tasks.

We considered whether the communicative goal of the response could be achieved even if the words that the user had added were not included in the message. For example, responding to a server’s request of what would you like to drink with water, just water please, or a glass of water please would still result in the server bringing the user water to drink. No significant differences were found between the communication support provided by each of the algorithms as measured by a two-way ANOVA ($F_{3,255} = 1.50$, $p = 0.21$, $\eta^2 = 0.01$).

A two-way ANOVA was used to test for differences between the combined vocabularies and the base vocabulary, which contains items that are used across contexts. This was done because the algorithms are intended to be used alongside this base vocabulary. Unfortunately, the effect size was too small for the difference to be measured by this test ($F_{4,319} = 1.76$, $p = 0.14$, $\eta^2 < 0.01$). As a result, a series of paired t-tests with Bonferroni correction were applied to compare each algorithm-augmented vocabulary to the base vocabulary. All of the combined vocabularies required significantly fewer additional words to support participant communication than the base vocabulary did on its own (Table 10). Participants also used a smaller proportion of the base vocabulary’s items than those that had been provided by the algorithms (Table 11). These two pieces of evidence demonstrate that the algorithms provided better support for communication than the base vocabulary did on its own.
Table 10 The average number of missing words and T-test results comparing the number of missing words from the combined vocabularies to that of the base vocabulary (\(M = 5.07, \ SD = 2.35\)).

<table>
<thead>
<tr>
<th>Combined Vocabulary</th>
<th>(M)</th>
<th>(SD)</th>
<th>(t(63))</th>
<th>(p^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Sites*</td>
<td>4.00</td>
<td>2.65</td>
<td>2.66</td>
<td>0.020</td>
</tr>
<tr>
<td>Dictionary*</td>
<td>3.89</td>
<td>2.30</td>
<td>3.58</td>
<td>0.001</td>
</tr>
<tr>
<td>Website*</td>
<td>4.14</td>
<td>2.61</td>
<td>2.48</td>
<td>0.031</td>
</tr>
<tr>
<td>Wikipedia-USF*</td>
<td>4.13</td>
<td>2.92</td>
<td>2.42</td>
<td>0.037</td>
</tr>
</tbody>
</table>

\(^1\) Bonferroni corrected
* Significantly different from the base vocabulary

4.3.1.4 Discussion

A vocabulary that had been created by humans did not offer more support than those generated by the algorithms. However, this base vocabulary did provide general communication support. Furthermore, when the vocabulary that was generated by the algorithms was combined with the base, it provided more support than the base could on its own. This combination provided participants with the vocabulary that they needed to compose their intended message and indicates that all four algorithms show the promise of supporting communication when combined with the base vocabulary.

If we look at the proportion of each vocabulary that was used to fulfill the participants’ communicative goals (Table 11), we can see that the generated vocabularies contain far fewer entries and that a higher proportion of their entries were used when responding to the location and theme based scenarios.

Even though the base vocabulary had many entries that were not needed by participants, it was able to support communication because it provided more expressions of politeness and phrases

Table 11 The average size of each vocabulary (in number of words) and proportion of each vocabulary that was used across all themes and locations (the base vocabulary was static).

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Average Size [as (M \ (SD))]</th>
<th>Average % Used [as (M \ (SD))]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Sites</td>
<td>20.3 (6.8)</td>
<td>30.6 (10.8)</td>
</tr>
<tr>
<td>Dictionary</td>
<td>30.0 (16.8)</td>
<td>22.0 (8.1)</td>
</tr>
<tr>
<td>Website</td>
<td>31.8 (2.9)</td>
<td>21.6 (6.6)</td>
</tr>
<tr>
<td>Wikipedia-USF</td>
<td>34.0 (0)</td>
<td>21.3 (8.1)</td>
</tr>
<tr>
<td>Base</td>
<td>209.0 (0)</td>
<td>4.1 (1.8)</td>
</tr>
</tbody>
</table>
that could be used in various situations. In general, it supported the participants who chose to be polite or friendly by providing the user with greetings, such as *hello*, politeness markers, such as *thank you*, and stereotypical phrases. Some examples of the stereotypical phrases that it supplied include *Do you have* and *I would like*. These phrases enable people to make requests and they were not provided by the output of the algorithms. Moreover, phrases of this nature should not be included in context-specific vocabularies because they are general expressions that could be used in almost any context. Users can access these all-purpose phrases and expressions of politeness with relative ease through the system’s search and navigation functions, making their use in a base vocabulary ideal since they can later be augmented with specific support.

Similarly, the developed algorithms provided poor support for confirmation and negation activities, whereas the base vocabulary supported this type of communication by including entries for *yes* and *no*. The base vocabulary also included the numbers one through four, which had not been included in the specialized vocabularies but were helpful to people who were trying to make purchases. Again, the general nature of numbers, agreement, and disagreement suggests that the core vocabulary for these functions belongs in a base vocabulary that can be built upon to provide more specialized support where a need arises.

The base vocabulary’s support for these simple communicative tasks reinforces the importance of augmenting it with other (more specific) vocabularies. This is justified further by the improved support that was provided when the output from each of the algorithms was combined with the base vocabulary. Combining the base vocabulary with the generated vocabulary enabled participants to communicate their messages politely.

An analysis of the responses of those who achieved their communicative goal, using the words in the combined vocabulary, revealed commonalities in the style of their replies: these participants tended to use short unembellished phrases. Those whose communication was well supported and who were polite also tended to use simple greetings and politeness markers, such as *thank you* and *hello* rather than indirect requests. Examples of the types of messages that could be supported even when words were missing can be seen in Table 8.

Indirect messages were supported poorly and could lead to confusion when key words were missing. For example, participant 2 wanted to respond with “just browsing, thanks” to refuse a
salesperson’s offer of assistance but just and browsing were not provided by the vocabulary. The resulting message of thanks would communicate the opposite intent by indicating gratitude for an offer of assistance. This implies that vocabularies should be adjusted when they are meant to support people whose culture or natural conversational style is less direct.

4.3.1.5 Conclusion

With respect to incorporating these techniques into a system, a subset of the algorithm-corpus pairs were chosen. The Review Sites and Dictionary algorithms were integrated because they provided similar levels of communication support while keeping the user’s search space to a minimum by consistently generating the smallest vocabulary collections. The dictionary algorithm was also chosen because we know that a collection of vocabulary items can always be generated due to the nature of the corpora over which the algorithm reasons.

The above algorithms only create a list of words. The resulting word lists provide no indication of the meaning of a word or phrase even though they can support communication. This means that people must be familiar with the items that are found in the generated list. Unfortunately, these lists of words and phrases provide inadequate support for ELLs. As a result, the next section reports on a study that aims to validate the use of different corpora for scaffolding ELL comprehension of the items from the lists.

4.3.2 Image Retrieval and Matching

This study aimed to validate the use of web-based corpora for providing visual representations of a word or phrase’s meaning. An Internet-based questionnaire was used to evaluate the images that were retrieved from 4 web-based corpora (see Table 12). These corpora were selected because of their breadth of vocabulary coverage and human-edited content. The study aimed to determine which corpora sufficiently communicated the meaning of individual vocabulary items.

4.3.2.1 Instruments and Procedures

Four corpora were evaluated for a subset of nouns and verbs that were selected based on their frequency of use within the English language. This list included 879 headwords. Care was taken to ensure that singular (738 words) and plural nouns (724 words), abstract (358 words) and
Table 12 The image corpora that were evaluated.

<table>
<thead>
<tr>
<th>Name</th>
<th>Corpus</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDClipArt</td>
<td><a href="http://www.pdclipart.org/">http://www.pdclipart.org/</a></td>
<td>This is a collection of public domain clip art where all of the images are drawn and have a cartoon-like feel. It has over 25,000 images and continues to grow.</td>
</tr>
<tr>
<td>Image-Net</td>
<td><a href="http://www.image-net.org/">http://www.image-net.org/</a></td>
<td>A collection of images that are organized using the WordNet hierarchy. Images are typically pictures and images continue to be associated with synsets. Images are not necessarily public domain.</td>
</tr>
<tr>
<td>CAPL</td>
<td><a href="http://capl.washjeff.edu/">http://capl.washjeff.edu/</a></td>
<td>This is a collection of images that are community generated and curated. All of the images are photographs of something in a culturally authentic context. The images are under a creative commons license (BY-NC-SA).</td>
</tr>
<tr>
<td>ESL Site</td>
<td><a href="http://www.eslsite.com">http://www.eslsite.com</a></td>
<td>This collection of cartoon-like images was assembled to support teachers of English as an additional language. Images are allowed to be used to support learning.</td>
</tr>
</tbody>
</table>

concrete nouns (354 words), and transitive (76 words) and intransitive verbs (58 words) were included in this set.

Convenience sampling of proficient English speakers was performed using a snowball approach to recruitment via email and social networking sites. Participants were asked to rate how well an image matched the meaning of a word on a 5-point Likert scale (1 - Strongly Disagree, 5 – Strongly Agree). The order of stimulus presentation was randomized based on the corpus.

Participants were asked to rate a set of up to 11 image-word pairs at a time but could rate as few as 8 items. After completing their ratings for a set, participants were asked if they wanted to rate more items. If they chose to continue participating, another set of stimuli was randomly selected. The stimuli were presented via a webpage (see Figure 9). This is where participants performed their ratings. No remuneration was provided but summaries of the results have been made available to the 6 participants who requested them.

4.3.2.2 Participants

Two-hundred and two English-language speakers performed 4754 ratings. The median number of ratings per participant was 16 items. This translates to participants rating two full sets of image-word pairs. The distribution of the number of participant ratings is skewed and can be seen in Figure 10.
English was the mother tongue of 75.61 percent of participants. The most common other mother tongues were French (9), Russian (5), Farsi (4), Korean (3), and Filipino (3). Participants had an average age of 39.7 years ($SD = 18.7$).

4.3.2.3 Results

Paired t-tests were used when participants had rated images from multiple corpora, and independent t-tests were used when participants had only rated images from a single corpus; Bonferroni correction was applied to prevent type 1 errors. The results of these tests are shown in Table 13; corpus pairs where there were no significant differences are not reported.

Participant ratings indicated that images from PDClipArt ($M = 3.42, SD = 0.65$) better communicated vocabulary item meaning than those from Image-Net ($M = 2.90, SD = 0.52$). The ESL corpus ($M = 3.42, SD = 0.44$) also outperformed Image-Net ($M = 2.90, SD = 0.52$) and CAPL ($M = 3.02, SD = 0.47$).
4.3.2.4 Discussion and Conclusion

The higher ratings that were received by the corpora that are drawing-based rather than photographic may be related to the abstract nature of drawings and people expecting a less precise representation when a sketch is provided. Regardless of the reason for the ability of sketch-based corpora to better communicate the meaning of individual words, it would appear that the PDClipArt and ESL corpora were only sufficient for communicating the meaning of a word.

Table 13 T-Test results comparing Corpus A to Corpus B for participant ratings of the appropriateness of an image for representing the meaning of a word.

<table>
<thead>
<tr>
<th>Corpus A</th>
<th>Corpus B</th>
<th>Independent T-Test</th>
<th>Paired T-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDClipArt</td>
<td>Image-Net</td>
<td>$t(45) = 3.54, \ p &lt; 0.01$</td>
<td>$t(30) = 3.85, \ p &lt; 0.01$</td>
</tr>
<tr>
<td>PDClipArt</td>
<td>CAPL</td>
<td>$t(43) = 1.39, \ p = 0.52$</td>
<td>$t(26) = 3.92, \ p &lt; 0.01$</td>
</tr>
<tr>
<td>ESL Verbs</td>
<td>Image-Net</td>
<td>$t(58) = 4.91, \ p &lt; 0.01$</td>
<td>$t(34) = 4.87, \ p &lt; 0.01$</td>
</tr>
<tr>
<td>ESL Verbs</td>
<td>CAPL</td>
<td>$t(53) = 2.51, \ p = 0.05$</td>
<td>$t(30) = 5.13, \ p &lt; 0.01$</td>
</tr>
</tbody>
</table>
The PDClipArt and ESL corpora were, therefore, integrated into the content recommendation that is part of the “static” language model that affects the interpret data and apply adaptation layers of the system.

4.4 Conclusion: Feature Design and Validation

The design guidelines that resulted from a proof-of-concept study of ELL use of a mobile communication-support tool were used to design a new MALL tool called VocabNomad. This proof-of-concept study had revealed several barriers to the use of cognitive and communication support tools by ELLs. The most prominent of which was the lack of support for emergent learner needs. The development and validation of content recommendation to support the emergent communication and educational needs of ELLs was then performed. A new tool was designed and the validated content recommendation mechanisms were incorporated.

The successful use of MyVoice by ELLs to support their language-learning activities and their communication provides evidence towards the validity of the newly designed system and the MALL approach that it aims to support. VocabNomad was designed to keep the beneficial features of MyVoice and improve upon its weaknesses. One of the primary improvements was the integration of on-demand support for communication and learning (DR2). The validation of this feature through two studies and the use of design-based evaluation techniques provides evidence towards the validity of using this approach to support ELL communication. The validation approach that was used was consistent with the layered approach to evaluating interactive adaptive systems.

Now that the individual system components have been validated to ensure that they are usable and that they will meet learner needs, the architecture of the system that is being used to demonstrate this approach to supporting ELLs can be described (Chapter 5) before discussing the validation of the tool as a whole (Chapter 6) and the subsequent evaluations of its use by learners (Chapter 7 and Chapter 8).
Chapter 5 - VocabNomad: System Description

Now that we know how VocabNomad has come to be (Chapter 3 and Chapter 4) we can discuss the system’s details. It is an adaptive dual-platform application (web and smartphone). I first illustrate the tool’s intended uses through user personae (5.1 Expected Uses). VocabNomad’s architecture is then described by highlighting the key elements of the layered approach upon which its design depends. The layered approach should not be confused with the more traditional multi-layered architecture that specifies a presentation, logic, and data layer. For the purposes of this chapter, it is enough to know that the layered approach decomposes a system into functional components that contribute to the five stages of reasoning (i.e., collect input data, interpret data, model the current state of the world, decide upon adaptation, and apply adaptation) that take place in adaptive systems. Additional details about the layered approach, its elements, and the relationship between them is provided in Chapter 4 since system development and validation were informed by the layered approach.

5.1 Expected Uses

The manner in which ELLs are expected to use VocabNomad is illustrated below using personae. These personae are based on participant reports and logs of system usage that were collected as part of the studies that are described in Chapter 3, Chapter 6, Chapter 7, and Chapter 8. None of the personae represent a single participant. They are representative amalgamations of different participant characteristics and behaviours (Y. Rogers et al., 2011). As personae, they depict general types of users through their description of a particular learner.

5.1.1 User Personae

Barbara is an undergraduate life sciences student visiting from Brazil. As part of her exchange program, she is participating in an academic English-language training program. Her English comprehension is strong but she struggles with composing her oral responses to others. She goes to English class every morning, rides the subway home with her roommate who is also a

---

6 VocabNomad was demonstrated at three conferences (Demmans Epp, Tsourounis, Djordjevic, & Baecker, 2013; Tsourounis, Demmans Epp, & Baecker, 2014; Mboutsiadis & Demmans Epp, 2015). See Appendix L for a complete list of publications and their relationship to my thesis work.
classmate, has lunch, and does some homework before going to the gym. She then returns home to watch television and visit with her roommates before she finishes her homework. About once a week she volunteers for a community organization where she gets to talk with people in English. She wishes to improve her English but finds that she often speaks with her roommates using Portuguese.

Jota is an engineering student who is visiting from Brazil. He watches television with his Canadian roommate who also serves as a source for seeking clarification of the English that Jota encounters. He only uses Portuguese when someone addresses him in Portuguese. Jota learned much of his English by reading technical materials and textbooks in English. While his course-based reading may have been done in English, he does some reading of Portuguese texts. Jota speaks easily and is comfortable asking for clarification, but lacks knowledge of Canadian pragmatics, especially with respect to politeness and formality.

Mohammad is a recent immigrant from Iran. He has a job as a construction worker and takes language courses as part of a government training program. He uses public transportation to get to and from work or school. Mohammad is strategically competent, outgoing, and friendly. He lacks vocabulary knowledge but makes up for it by using other strategies to communicate. He dislikes using his mother tongue and tries to use English whenever possible.

Tara is a Chinese housewife who now works in a factory making dumplings. She is taking English classes and is motivated to learn but lacks confidence. Tara has become frustrated by her apparent lack of progress. She struggles with understanding and producing English but has a strong desire to learn to communicate in English.

Koji is a Japanese high school student. He plays basketball and likes to talk with others. He is a conscientious student who listens to English music. Koji is willing to take the risk of talking with English speakers when he needs access to the knowledge or expertise that they possess or he can identify a common area of interest.

5.1.2 Personae Use of VocabNomad

Barbara uses VocabNomad to review vocabulary and to find the words that she needs while writing essays.
Jota comes across a word that he does not know while doing his course readings. He opens VocabNomad to record this word by touching the plus sign.

He creates a new vocabulary entry.

He types in graph and clicks on the camera. Because Jota has no photos, he is only given the option to Cancel or use the Camera. He clicks on Camera.

Jota takes a picture of one of the graphs in his textbook and approves it by clicking on the checkmark.

Jota then types a definition for a graph and uses the prediction to help him fix typos and spell words.

He then adds a tag to the word so that he can more easily find it later.

He saves his new word and is returned to the main screen where his word has been added to the list of all vocabulary entries.

**Figure 11** A storyboard showing the screens that Jota would see were he recording a newly encountered word.

Jota has an advanced command of English and is comfortable relying on the human-support that he has at home. As a result, he uses VocabNomad very little. His primary use of the system is to record the new language that he encounters so that he can review it later or ask his home-stay host about it. A specific example of his system usage can be seen in Figure 11.  

---

7 All of the examples of system usage are illustrated through screenshots of the application being used. However, I preformed the activities that were used to generate the screenshots.
Mohammad adds the words that he thinks are essential to VocabNomad. These are mostly the words that he commonly uses in his L1 (Farsi). He then studies those and other words while riding the bus or subway.

An example of Tara’s usage can be seen in Figure 12. She uses VocabNomad to help her get the foods that she wants when she goes to the grocery store. She scans through the pictures of Tara opens the application while she is at the grocery store.

She searches for fruit.

She scans through the fruits until she finds the type of melon that she wants to buy. She then asks the grocer where she can find them.

After returning home and cutting up her melon, Tara practices how to say honeydew. So, she listens to the pronunciation model before recording herself saying honeydew.

She then plays her recording before listening to the system pronounce honeydew. Tara continues to practice until she thinks her pronunciation matches that provided by the system.

Figure 12 A storyboard showing the screens that Tara would encounter were she trying to learn how to say honeydew.
vocabulary items for food until she finds the name of the food that she wants so that she can ask someone where to find it. She also uses VocabNomad to confirm her knowledge before speaking to someone and will show the picture to the interlocutor when communication is especially difficult. She will sometimes study and practice her pronunciation when she is alone or waiting on others.

Koji uses VocabNomad to study for his final exam. He will occasionally modify vocabulary sentences so that they are more meaningful for him and he adds new words to the system so that his classmates can learn them if they want to. Koji’s use can be seen in Figure 13.

---

**Figure 13** A storyboard showing the screens that Koji would encounter when studying or editing a word.
5.2 User Interface

To support these users, VocabNomad employs graphical user interfaces on two platforms: web browsers and android mobile phones. Each interface provides slightly different functionality but the symbols that are used to communicate information are consistent across platforms. Both interfaces allow the user to view, modify, or add learning content; search through learning content; or request new learning content. Given the conventions of the platforms and the constraints they place on the interface, each user interface will be described separately.

5.2.1 Mobile Interface

Screen real estate is more constrained on mobile devices than it is on desktop or laptop computers. This resulted in summary information for vocabulary items being presented on the main application screen and more detailed information being accessible on demand (Figure 14). This detailed view is where users can listen to audio versions of the different components of a vocabulary entry or where they can record authentic language use and associate it with a vocabulary entry (Figure 14). The example shown through Figure 14 demonstrates how the usage example does not necessarily have to match the definition that is provided through VocabNomad. While this may cause confusion for some learners, the usage example and image still provide the learner with a sense of one of the valid definitions of test. This type of inconsistency is likely to arise when learners modify examples so that they are salient to their current situation.

Figure 14 The vocabulary list (left), a single word (centre), and its edit screen (right).
The ability to record interesting language usage through the mobile interface was added because the device easily affords the recording and playback of audio materials. Likewise, users can photograph examples of vocabulary items that they encounter in everyday settings. This device affordance allows the application to better support the learners’ needs for noticing and recording language use from their surrounding environment.

To support emergent learner needs, the mobile application enables the creation of new vocabulary collections and the translation of words from the learner’s mother tongue to English using Microsoft Translator (Microsoft Translator Team, 2013). This functionality is provided through the same interface as that used for conducting searches (Figure 15).

5.2.2 Web Interface

The web interface displays all vocabulary information on the primary screen, which is searchable (Figure 16). It allows users to browse through collections of learning materials and choose ones that they wish to import so that they can study the materials and prepare themselves for any task that they foresee themselves having to do (Figure 17). The web interface allows users to complete vocabulary tests (Figure 18) and retrieve learning materials that they may have deleted by accident (Figure 19). The web-interface also permits users to add learning materials or edit existing vocabulary (Figure 20).
Figure 16 The Vocabulary list view of the web interface following a search for milk.

Figure 17 The page where users can import collections of vocabulary items.
Figure 18 The test page. This test has been configured to include translation and sentence production tasks.

Figure 19 The page where users can see which items they have deleted and retrieve them.
5.3 System Architecture

The key features and system elements that enable the VocabNomad uses that are illustrated through the above-described personae are detailed below. VocabNomad was developed using a variety of tools. The mobile client was developed in Java with an SQLite database. The web interface was developed with ASP.NET, C#.NET, CSS, and xhtml. The server-side functionality uses xml, C#.NET, python, the NLTK, Elmah error reporting, and a SQlServer database. Communication between the server and android client is handled using RESTful webservices (Richardson & Ruby, 2007) that transmit JSON objects. Automated testing of the web-based interface was performed using the Selenium IDE (Huggins et al., 2011).

Figure 21 shows where some of the architectural elements fit within the layered approach. It highlights how the boundaries between collecting and interpreting data (the green wedges) are
Figure 21 VocabNomad’s architecture as seen through the lens of the layered approach.

less pronounced in this system. The same is true of the boundaries between modeling the current state of the world and the decide upon adaptation layer (the purple wedges).

5.3.1 Domain Model

If we want to look at the domain model from a learning objects perspective, where learning content is encapsulated and reusable (Churchill, 2006), we can see a hierarchy of learning object types and granularities within VocabNomad. At the top level, there are categories of content: these are the tags that are associated with vocabulary entries. At the intermediate level, we have vocabulary entries themselves, and at the bottom level we have the images that can be associated with individual vocabulary entries. The finest grained type of learning object is the image. It can be reused by individual vocabulary entries, as is the case for bike and bicycle (Figure 22). Vocabulary entries can be associated with one or many tags, which allows them to be reused across contexts. The largest of the learning objects are the collections of vocabulary entries that are defined by the association between vocabulary entries and tags where each tag represents a
learning object. These learning objects are reusable in the sense that they can be shared by learners so that other learners can benefit from the vocabulary creation efforts of their peers.

The content that is contained within these learning objects, the underlying pronunciation model from the on-board android text-to-speech engine, and the learning objects themselves make up the system’s domain model. This domain model is a representation of English vocabulary and its usage. As such, it is graph-based to allow for greater flexibility and the reuse of items rather than the more typical hierarchical approach that is found in ontologies. This allows for greater ambiguity in the description of the domain. However, it holds the potential to reflect the ambiguity that surrounds the ever-evolving and shifting usage of natural languages.

The domain model is represented through the following tables and views: vocab, partsOfSpeech, synonyms, tags, languageMapping, language, vocabPOS, vocabTagPairs, VocabView, vocabPOSOView, and EngWordFreq. The domain content is described by the vocab, tags, language, languageMapping, and EngWordFreq tables. Domain metadata is detailed through the language and partsOfSpeech tables. The relationships between individual vocabulary entries are described by synonyms, vocabPOS, and vocabTagPairs. These relationships are fully expressed in the VocabView and VocabPOSView. Additional information about the role of each of the database entities within the domain model is provided in Table 14. An entity-relationship diagram for the system’s database can be seen in Appendix G.

The content that is contained within the domain model indirectly captures information about English varieties and represents this information through the sentences that show a word being used in context. However, the system does not specifically draw the learner’s attention to this
information. This content was created using three methods: developer-generation, system-
generation, and user-generation. The developer-generated content follows a curated model of
content creation. As such, that content is relatively static and only changes when a specific need
arises. Collections of vocabulary are added as time permits. System-generated content is created
from Internet-based resources using information retrieval techniques where the system generates
content as the result of an unmet or emergent need that the user has identified. User-generated
content stems from a collaborative approach where individual users manually create and share
content through the addition of new content or by editing existing content. In this way, the
domain model is continually updated through the creation and modification of the user and
system-generated content. This approach to content development also means that content quality
and the richness of the provided usage examples may vary. However, the continual review and
editing that users can perform means that the content quality is improved over time. Details
about these approaches to content generation are provided below.

Table 14 A subset of the different database entities and their role within the system’s
domain model.

<table>
<thead>
<tr>
<th>Database Entity</th>
<th>Description</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>vocab</td>
<td>contains individual vocabulary items and their associated content</td>
<td>it is the primary source of data for the domain model and learning content</td>
</tr>
<tr>
<td>partsOfSpeech</td>
<td>categories of word types</td>
<td>metadata: grammatical information</td>
</tr>
<tr>
<td>synonyms</td>
<td>relationship between different vocabulary entities</td>
<td>metadata: semantic relationships between different learning content</td>
</tr>
<tr>
<td>tags</td>
<td>category names</td>
<td>this is part of the graph-based domain model and a primary source of data for the domain model and learning content</td>
</tr>
<tr>
<td>language</td>
<td>language names</td>
<td>metadata: semantic information</td>
</tr>
<tr>
<td>languageMapping</td>
<td>mapping of vocabulary items from L1 to L2</td>
<td>used to ensure consistency in learner attributes</td>
</tr>
<tr>
<td>vocabPOS</td>
<td>the relationship between parts of speech and vocabulary entries</td>
<td>metadata: relationships between vocabulary and grammatical information</td>
</tr>
<tr>
<td>vocabTagPairs</td>
<td>provides the mapping between vocabulary items and tags</td>
<td>metadata: semantic links between vocabulary items based on their tags (categories)</td>
</tr>
<tr>
<td>VocabView</td>
<td>most of the information about vocabulary</td>
<td>joins the different types of learning objects together for use in the system</td>
</tr>
<tr>
<td>VocabPOSView</td>
<td>combines the part of speech information with the VocabView</td>
<td>it is not currently used but could be used to enable additional adaptivity and instruction</td>
</tr>
<tr>
<td>EngWordFreq</td>
<td>the 5000 most commonly used English words and their frequency of use</td>
<td>used for adaptive testing</td>
</tr>
</tbody>
</table>
5.3.1.1 Developer-Generated Content

This curated collection of learning content contains two parts. The first is the default vocabulary that is added to all user accounts. The second is a set of learning content collections that users can import. These collections were manually created by the development team.

The initial list of default vocabulary was based on that used in the aphasia deployments of MyVoice. This collection was developed in partnership with a teacher and a speech language pathologist. The 1000 most commonly occurring English words, as identified within the British National Corpus and the Corpus of Contemporary American English (Davies & Gardner, 2010; Leech, Rayson, & Wilson, 2001) were then added to this list. Additional sets of vocabulary items that related to food, beverages, numbers, weather, and colours were added in an opportunistic manner with some consideration for the types of topics that people would typically encounter in their everyday lives. Additional categories of learning content were occasionally added. These collections included words related to human anatomy, animals, clothing, cooking, and furniture.

All of the vocabulary items that were added had tags associated with them. Where it was possible, images were found and associated with each vocabulary entry. Sentences and definitions were added to the vocabulary entries. All of the curated content was required to have an image or a sentence, with a preference for each vocabulary entry (i.e., learning object) to have both a sentence and an image associated with it. The sentences were either taken from the WordNet corpus of the NLTK (Princeton University, 2010), in which case they may have been modified, or they were written by the content developers. Definitions were prepared by consulting dictionaries, by copying or modifying those found in WordNet, or by the content developers. Regardless of the manner in which the sentences and definitions were prepared, they were later verified by the lead developer. The images that were used were selected from the Noun Project and other open source icon projects, found using Google image search, or created by content developers. Like the definitions, synonym associations between words were developed by consulting dictionaries and later reviewed by the lead developer.

5.3.1.2 System-Generated Content

Algorithm-generated content is created when a user requests vocabulary for a topic that does not have learning content associated with it. In this case, a tag that matches the user’s search term is
created and a two part process is initiated. The first part of this process generates a collection of vocabulary items (i.e., words and phrases) and the second part finds images to associate with individual vocabulary items.

The user initiates this process by conducting a tag search for a category of words that the user does not already have associated with his or her account. The pseudo-code from Figure 23 is then run. The algorithms that were validated in Chapter 4 are only called when an existing collection that meets the user’s identified need cannot be found. Once the results of the vocabulary generation algorithms have been combined, they are passed into the image retrieval method so that appropriate images can be identified.

5.3.1.3 User-Generated Content

The other way in which users can generate content is by explicitly adding vocabulary items and associating those items with a category or tag. This will add the item to that user’s account. Vocabulary entries that have tags associated with them and that have been shared by a user can be imported by other users.

This import process can be initiated through a search that makes a request to the system’s on-demand support. The imported materials are then added to the user’s account if they already exist

```plaintext
OnDemandVocabSupport(topic)
    if(!importableContent.Has(topic))
        return topic
    else
        a[] <- GenerateVocab(algA)
        b[] <- GenerateVocab(algB)
        result[] <- merge(a, b)
        ImageRetrieval(result)
        return result

ImageRetrieval(words)
    foreach(item w in words)
        img <- search(corpusA)
        if (img not null)
            w.img <- img
        else
            img <- search(corpusB)
            if(img not null)
                w.img <- img
        img <- null
        return
```

Figure 23 The pseudo code for the on-demand vocabulary generation feature.
or they are generated using the above described algorithms when they do not. The user can edit any of these entries and share them with the larger VocabNomad learning community. In this way, existing learning materials have the potential to be continually improved.

### 5.3.2 Model of the World

Some of the system’s model of the world is encapsulated within the learner model attributes. This includes the tracking of when and where certain vocabulary is used as well as the tracking of the locations that are most frequented by learners. This data is held within the following tables: `userEvents`, `location`, `activities`, `activityTagPairs`, and `Time`.

At present this data is only being tracked. However, the system has architectural support for the future adaptation of learning materials based on the information that can be inferred about the current state of the learner’s world using this information.

### 5.3.3 Learner Model

The ecological approach to learner modeling was used because it can support predetermined and emergent needs (McCalla, 2004). It was also used because it enables the detailed tracking of ELL interactions with learning materials. This allows us to better explore which user actions are related to learning outcomes (Ma, 2013). However, this choice in modeling technique requires the re-envisioning of the learner model as a process rather than an object.

The ecological approach divides the learner model into two components: long-term learner attributes and the learner’s episodic history (McCalla, 2004). This episodic history details the learner’s interactions with learning resources, which includes learning objects but could include peers or instructors. This section details the data upon which the learner model is based, the modeling process that is used, and the adaptive theory upon which the system acts.

#### 5.3.3.1 Data

The data that are part of the learner model consist of the user’s characteristics, which include his or her mother tongue, the user’s role (learner or teacher), the locations that the user frequently visits, and the activities that she or he performs regularly (see Table 15). These are conceptually part of the learner’s characteristics, but not all of them are required and this data is stored across
Table 15 The learner’s characteristics portion of the learner model. These data describe the user’s more stable characteristics and are spread throughout the users, UserRole, UserCourses, activities, location, and Time tables in the database. They are presented jointly to facilitate comprehension of the model attributes rather than to provide details about the model’s implementation.

<table>
<thead>
<tr>
<th>User</th>
<th>L1</th>
<th>Role¹</th>
<th>Course¹</th>
<th>Activities²</th>
<th>Locations²</th>
<th>Time²</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>shopping, working, studying</td>
<td>home, work, school, ...</td>
<td>morning, noon, lunch, afternoon ...</td>
</tr>
<tr>
<td>import</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>shopping, working, studying</td>
<td>home, work, school, ...</td>
<td>morning, noon, lunch, afternoon ...</td>
</tr>
<tr>
<td>Nihon_S01</td>
<td>Japanese</td>
<td>Student</td>
<td>JapanHS</td>
<td>shopping, working, studying</td>
<td>home, work, school, ...</td>
<td>morning, noon, lunch, afternoon ...</td>
</tr>
<tr>
<td>Nihon_S02</td>
<td>Japanese</td>
<td>Student</td>
<td>JapanHS</td>
<td>shopping, working, studying</td>
<td>home, work, school, ...</td>
<td>morning, noon, lunch, afternoon ...</td>
</tr>
<tr>
<td>Nihon_S03</td>
<td>Japanese</td>
<td>Student</td>
<td>JapanHS</td>
<td>shopping, working, studying</td>
<td>home, work, school, ...</td>
<td>morning, noon, lunch, afternoon ...</td>
</tr>
<tr>
<td>Sensei_01</td>
<td>Japanese</td>
<td>Teacher</td>
<td>JapanHS</td>
<td>shopping, working, studying</td>
<td>home, work, school, ...</td>
<td>morning, noon, lunch, afternoon ...</td>
</tr>
</tbody>
</table>

¹ Users can only have 1 role but can be in multiple courses.
² Users can have multiple activities, locations, and times. The specification of location (GPS information) and time attributes (hours of the day) can be set to user-specific values, new times and locations can also be added for individual users. User-specific activities can be added but their basic definition is not adjustable.
³ Foreign key references were changed to match their associated values to increase readability and interpretability.

various tables. Only the learner’s mother tongue is required. The course that a learner is enrolled in is also required should the learner want to complete any of the adaptive tests.

The episodic history of the learner model consists of click-stream level logs of every interaction that a learner has had with the system (see Table 16). This logging includes learner interactions with individual vocabulary items, such as editing a sentence or having it read aloud. The logging

Table 16 The episodic history portion of the learner model¹.

<table>
<thead>
<tr>
<th>userID</th>
<th>eventTime</th>
<th>eventType</th>
<th>time</th>
<th>tag</th>
<th>activity</th>
<th>LangMapID</th>
<th>location</th>
<th>userEntry</th>
<th>longitude²</th>
<th>latitude²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESM01</td>
<td>2014-10-27</td>
<td>Add Voc: A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NULL</td>
<td>NULL</td>
<td>NULL</td>
<td>-79.397</td>
<td>43.659</td>
</tr>
<tr>
<td></td>
<td>16:51:18.710</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESM02</td>
<td>2014-10-27</td>
<td>L2 Search:A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7-Up</td>
<td>7-Up</td>
<td>NULL</td>
<td>-79.397</td>
<td>43.659</td>
</tr>
<tr>
<td></td>
<td>16:12:24.737</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESM02</td>
<td>2014-10-27</td>
<td>L2 Search:A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>UV index</td>
<td>UV index</td>
<td>NULL</td>
<td>-79.397</td>
<td>43.659</td>
</tr>
<tr>
<td></td>
<td>16:17:22.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ This is a human-readable version of the userEvents table where the foreign-key references have been replaced with their associated values and where columns that logged redundant information have been removed.
² The longitude and latitude have been truncated to give a sense of the data without revealing the actual locations of participants.
also includes learners’ requests for new learning materials and their information seeking practices (e.g., scrolling, translating words into English, or use of the search feature).

The complete list of learner actions that are tracked as part of a learner’s episodic history and the

**Table 17 A listing and description of the event types that are logged by the web interface.**

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Description</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test:W</td>
<td>User submitted an adaptive test</td>
<td></td>
</tr>
<tr>
<td>L1 search:W</td>
<td>User searched for a word in L1 through web interface</td>
<td></td>
</tr>
<tr>
<td>L2 search:W</td>
<td>User searched for a word in L2 through web interface</td>
<td></td>
</tr>
<tr>
<td>Tag Search:W</td>
<td>User searched tags through web interface</td>
<td></td>
</tr>
<tr>
<td>Add Voc:W</td>
<td>User initiated the addition of a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Import Voc:W</td>
<td>User imported a set of vocabulary items through web interface</td>
<td></td>
</tr>
<tr>
<td>Edit Voc:W</td>
<td>User initiated the editing of a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Delete Voc:W</td>
<td>User deleted a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Restore Voc:W</td>
<td>User restored (undeleted) a vocabulary item through web interface</td>
<td></td>
</tr>
<tr>
<td>View Voc:W</td>
<td>User viewed list of vocabulary items through web interface</td>
<td></td>
</tr>
<tr>
<td>View Deleted Voc:W</td>
<td>User viewed list of deleted vocabulary items through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Login:W</td>
<td>User logged in through web interface</td>
<td></td>
</tr>
<tr>
<td>Logout:W</td>
<td>User logged out through web interface</td>
<td></td>
</tr>
<tr>
<td>Add tags:W</td>
<td>User added tags to vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Edit Tags:W</td>
<td>User edited the tags associated with a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Delete Tags:W</td>
<td>User removed the tags associated with a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Cancel Voc Edit:W</td>
<td>User cancelled their vocabulary item editing through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Save Voc Edit:W</td>
<td>User saved their changes to the vocabulary item during edit through web interface</td>
<td>✓ ✓</td>
</tr>
<tr>
<td>Add Another Voc:W</td>
<td>User initiated the addition of another vocabulary item from the add/edit vocabulary page through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Share:W</td>
<td>User shared a vocabulary item through web interface</td>
<td></td>
</tr>
<tr>
<td>Unshare:W</td>
<td>User unshared a vocabulary item through web interface</td>
<td></td>
</tr>
<tr>
<td>Img Edit:W</td>
<td>User edited the image associated with a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Word Edit:W</td>
<td>User edited the word field of a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Sentence Edit:W</td>
<td>User edited the sentence field of a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Defn Edit:W</td>
<td>User edited the definition field of a vocabulary item through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Next Page:W</td>
<td>User went to next page in list of current/deleted vocabulary items through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Prev Page:W</td>
<td>User went to previous page in list of current/deleted vocabulary items through web interface</td>
<td>✓</td>
</tr>
<tr>
<td>Data Sync:W</td>
<td>Data synchronization was performed between the client and server</td>
<td></td>
</tr>
<tr>
<td>Request Voc:W</td>
<td>User requested vocabulary through web interface</td>
<td></td>
</tr>
</tbody>
</table>
learning content with which she or he interacts is detailed in Table 17 and Table 18. New events
can be added to this list whenever new features are added; the addition of new types of logging
events requires that an SQL script be run. It also requires that both the server and android
applications are restarted. The Syn. and Test columns of Table 17 and Table 18 specify whether
that type of log event is used as part of the reasoning for the adaptive synonym recommendation
or the selection of items for the adaptive test feature.

5.3.3.2 Modeling Process

Rather than having a model that is updated, as is the case with most learner modeling
approaches, VocabNomad uses a dynamic modeling process. It reasons over the data contained

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Description</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Sync:A</td>
<td>Data synchronization was initiated by the android client</td>
<td></td>
</tr>
<tr>
<td>L1 search:A</td>
<td>Use searched by L1 through android client</td>
<td></td>
</tr>
<tr>
<td>L2 search:A</td>
<td>User searched words by L2 through android client</td>
<td></td>
</tr>
<tr>
<td>Tag Search:A</td>
<td>User searched tags through android client</td>
<td></td>
</tr>
<tr>
<td>Add Voc:A</td>
<td>User added a vocabulary item through android client</td>
<td>✓✓</td>
</tr>
<tr>
<td>Edit Voc:A</td>
<td>User edited a vocabulary item through android client</td>
<td>✓✓</td>
</tr>
<tr>
<td>Delete Voc:A</td>
<td>User deleted a vocabulary entry through android client</td>
<td>✓</td>
</tr>
<tr>
<td>View Voc:A</td>
<td>User viewed vocabulary list through android client</td>
<td>✓</td>
</tr>
<tr>
<td>View Detailed</td>
<td>User viewed a vocabulary item in detail through android client</td>
<td>✓✓</td>
</tr>
<tr>
<td>Voc:A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Login:A</td>
<td>User logged in through android client</td>
<td></td>
</tr>
<tr>
<td>Add tags:A</td>
<td>User added tags to a vocabulary item through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Edit Tags:A</td>
<td>User edited the tags associated with a vocabulary item through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Delete Tags:A</td>
<td>User deleted tags associated with a vocabulary item through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Cancel Voc Edit:A</td>
<td>User cancelled vocabulary item editing through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Save Voc Edit:A</td>
<td>User saved changes to a vocabulary item through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Share:A</td>
<td>User shared a vocabulary entry through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Unshare:A</td>
<td>User unshared a vocabulary entry through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Img Edit:A</td>
<td>User edited the image associated with a vocabulary entry through android client</td>
<td>✓</td>
</tr>
<tr>
<td>JIT Img:A</td>
<td>User requested just-in-time image support through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Img Library: A</td>
<td>User selected image from library through android client</td>
<td></td>
</tr>
<tr>
<td>Img Photo: A</td>
<td>User took photo through android client</td>
<td></td>
</tr>
<tr>
<td>Word Edit:A</td>
<td>User initiated edit of vocabulary entry through android client</td>
<td>✓</td>
</tr>
<tr>
<td>Sentence Edit:A</td>
<td>User edited the sentence associated with a vocabulary entry through android client</td>
<td>✓</td>
</tr>
</tbody>
</table>
in the learner’s episodic history to infer vocabulary knowledge and recommend additional learning resources. It also uses the learner’s episodic history to dynamically select test items should the learner decide to test his or her knowledge of English vocabulary. Since the model is viewed as a process or function that dynamically applies a theory of adaptation, the next section provides additional details about the modeling process.

5.3.3.3 Adaptive Theory

VocabNomad currently provides two types of adaptivity: synonym recommendation and testing.

Synonym recommendation infers learner knowledge based on the number of times that she or he has viewed a particular vocabulary entry. This inference is based on theories of fast mapping (Carey, 2010; Wagner et al., 2007), extended mapping, and bootstrapping (Carey, 2004) that describe the acquisition of vocabulary based on exposure to their use in context. Figure 24 describes the states through which the system passes when inferring learner knowledge and determining which synonyms to display. It is worth noting that each vocabulary item has its own

![Figure 24](image)

**Figure 24** A state diagram that details the recommendation conditions for synonyms. \( VC_i \) is the view count for the specified vocabulary entry, \( S_i \) is the visibility of synonyms for vocabulary entry \( i \), and \( S_{si} \) is the synonym visibility of \( i \)'s synonyms from the perspective of vocabulary item \( i \) only. \( S_{si} \) could be true if another one of its synonyms (e.g., \( j \)), is known (i.e, \( VC_j \geq 4 \)).
state; reasoning over learner knowledge and the accompanying recommendation of learning materials is done on a vocabulary entry by vocabulary entry basis.

Once the system has inferred that a learner knows a word, its near synonyms are displayed via the mobile interface. The near synonyms for the synonyms of a word for which the learner’s knowledge has been inferred are also displayed. This carry forward in adaptive recommendation stops when the synonym relationship exceeds 1 degree of separation. See Figure 25 for an example, where the star is the vocabulary entry that the learner model has inferred is learned, the hexagons represent the synonyms that will also display their synonyms, and the circles are synonyms that are displayed.

The second form of adaptivity is provided through the web-based testing feature (Figure 18). The test consists of any number of items provided that number is a multiple of four. For each multiple of four, four words from the 5000 most commonly used English words are randomly selected, with one word coming from each quartile based on the word’s frequency of use in English (see Appendix H for additional details about the word list).

Figure 25 Synonym recommendation propagation for SUV, with lines between entities indicating synonymous relationships. The system has inferred that the learner knows SUV (the star). Its first-degree connections (the hexagons: Sport Utility Vehicle and Vehicle) also display their synonyms, but second degree or higher connections do not display their synonyms (the circles: truck and lorry).
Table 19 Example adaptive tests that are based on the frequency of use of a word within English (English Freq. Based) or the frequency with which participants interacted with the vocabulary item (Usage Based). ESM01 and ESM02 are the participants who completed these tests.

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Usage Based</th>
<th>English Freq. Based</th>
<th>Usage Based</th>
<th>English Freq. Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Hardest</td>
<td>hummus</td>
<td>pendant</td>
<td>toss</td>
<td>cheeses</td>
</tr>
<tr>
<td></td>
<td>balaclava</td>
<td>ancho chili</td>
<td>file</td>
<td>salts</td>
</tr>
<tr>
<td></td>
<td>midori</td>
<td>tuna</td>
<td>chamber</td>
<td>cod</td>
</tr>
<tr>
<td></td>
<td>apples</td>
<td>Memphian</td>
<td>information</td>
<td>chicken</td>
</tr>
<tr>
<td>2</td>
<td>static</td>
<td>summon</td>
<td>everybody</td>
<td>woman</td>
</tr>
<tr>
<td></td>
<td>probability</td>
<td>manager</td>
<td>fortune</td>
<td>literacy</td>
</tr>
<tr>
<td></td>
<td>one thousand one hundred and five agent</td>
<td>ash</td>
<td>material</td>
<td>visual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tall</td>
<td>family</td>
<td>flurry</td>
</tr>
<tr>
<td>3</td>
<td>a million</td>
<td>pumpkin</td>
<td>produce</td>
<td>blackberries</td>
</tr>
<tr>
<td></td>
<td>freshman</td>
<td>barley</td>
<td>hunter</td>
<td>west</td>
</tr>
<tr>
<td></td>
<td>sophomore</td>
<td>guacamole</td>
<td>dilemma</td>
<td>sole</td>
</tr>
<tr>
<td></td>
<td>fourth</td>
<td>boa</td>
<td>so</td>
<td>bagel</td>
</tr>
<tr>
<td>4 - Easiest</td>
<td>cream</td>
<td>mushrooms</td>
<td>style</td>
<td>buns</td>
</tr>
<tr>
<td></td>
<td>macaroni</td>
<td>lime</td>
<td>hedge</td>
<td>banger</td>
</tr>
<tr>
<td></td>
<td>acorn squash</td>
<td>vinegars</td>
<td>student</td>
<td>man</td>
</tr>
<tr>
<td></td>
<td>bowl</td>
<td>a million</td>
<td>trial</td>
<td>baby</td>
</tr>
</tbody>
</table>

The adaptive testing can use information from the learner’s episodic history to randomly select 4 vocabulary items. This information is employed when the test is configured to allow for usage-based testing. In these cases, vocabulary items are selected based on the individual learner’s level of interaction with vocabulary items in the system. One test item comes from each quartile based on the frequency with which that learner has interacted with the vocabulary item. This inference is made over any of the events that are associated with vocabulary items. These events are indicated with a checkmark (✓) in the test column of Table 17 and Table 18. If the user has not interacted with enough items prior to the test when usage-based testing is enabled, the remaining test items are selected from the collection of commonly used English words. Example tests can be seen in Table 19: these examples are drawn from the tests that were generated during a study of VocabNomad’s relationship to learning, affect, and communication (Chapter 8). At present, these tests are marked by humans.
5.4 Conclusion

The above-described system was designed to be highly flexible within the context of providing support for communication and vocabulary learning. It was also designed to allow for and accommodate the many ambiguities that accompany language use while allowing the user to exercise a high-level of control.

The system allows users to request, create, modify, or share learning materials. Its use of the ecological approach to learner modeling enables the support that it provides to learners, via adaptive content recommendation, to be modified. This is possible because of the re-envisioning of the learner model as a dynamic process, and it means that changing how VocabNomad infers learner knowledge requires the adjustment of a function. In VocabNomad’s case, this means that an SQL query would need to be changed.

Having seen user personae, the validation of specific features, and a description of the system itself, it is now time to discuss the evaluation of this approach to MALL that has been operationalized through the described tool. Its evaluations aim to answer the extent to which tool use can support the communication and vocabulary learning activities of ELLs (RQ1), how ELLs integrate the system into their communication and vocabulary learning activities (RQ2), and the relationship that exists between VocabNomad usage and each of three aspects of language learning (vocabulary knowledge, affect, and communicative success) (RQ3). The answers to these questions come from studies that were conducted across educational and cultural contexts and can be seen in Chapter 6, Chapter 7, and Chapter 8. We will start with an exploration of the use of this communication and learning support tool by advanced ELLs who are living in an English-language dominant environment.
Chapter 6 - VocabNomad: System Validation Study

Recently there have been calls for MALL tools that are fundamentally different from previous MALL or CALL systems and the paper exercises on which they are often based (Ballance, 2013; Burston, 2014b). The VocabNomad system aims to meet these calls by supporting the communication and vocabulary acquisition of ELLs (Demmans Epp et al., 2013). Its design and development was user-centered and based on previous language-learning research as well as multiple evaluations of individual features, the system’s design, and its functionality. This development and the system are described in Chapter 3, Chapter 4, and Chapter 5.

A final formative evaluation of the system was performed to see if the newly designed system could support users’ learning activities as well as their communication⁸. It was important to ensure that these two activities were still supported following the system’s development since the support provided for these activities could have been broken at any point during the system design and refinement. This study provides a lower-stakes evaluation of the system as a whole before subjecting learners in formal educational environments to what could have been a useless tool. Additionally, this deployment was used to further validate the interaction history and learner model that belong to the dynamic models component of the system.

This final formative evaluation was performed following the validation of the new system’s user interface and the integration of the adaptive learning content recommendation. Like the original proof-of-concept study (Chapter 3), this evaluation aimed to explore ELL use of a mobile tool. However, the evaluation was performed using an application that was designed to address additional ELL needs. This included their need for support during unplanned events.

The evaluation collects evidence that is used to inform our understanding of how the tool can be integrated into ELL communication and learning practices (RQ2) as well as the extent to which it can support the vocabulary learning and communication needs of ELLs (RQ1).

---

⁸ Different aspects of this study appear in Tsourounis, Demmans Epp, and Baecker (2014), Demmans Epp (2014), and Tsourounis and Demmans Epp (Accepted). See Appendix L for a complete list of publications and their relationship to my thesis work.
6.1 Instruments and Procedures

Similar procedures to those of the previous exploratory study were followed (see 3.2 Methodology). However, a different mobile tool, VocabNomad, and an updated demographics form were used (see Appendix B). The interview protocol was modified to collect information about the types of feedback that ELLs desire (see Appendix C). However, this chapter will not report on that aspect of the study since it is outside of the scope of this thesis.

VocabNomad is similar to MyVoice in that it provides users with image-word pairs that can be verbalized using text to speech. VocabNomad gives users access to vocabulary item definitions and examples of the items being used in sentences. This tool also allows users to record language that they have encountered in their environment.

The deeper differences between these applications relate to how content is organized and recommended to users. VocabNomad uses a flat navigation structure that is searchable using English or the user’s mother tongue, and it allows the user to request content based on an identified need or interest. Learning and communication support materials are then generated using the algorithms and corpora that were validated in Chapter 4. VocabNomad tracks user actions within the system and builds a model of the user’s vocabulary knowledge. At present, the decision to expose the user to new vocabulary, by displaying synonyms for known words, is based on information contained in this model and the theory of bootstrapping or extended mapping. Application details are provided in Chapter 5.

Participants used VocabNomad for approximately three weeks, after which interviews were conducted. The interviews lasted from 27 to 77 minutes and were transcribed by the interviewer and one of his lab mates. These transcripts were then checked by myself before they were analyzed by the interviewer and me. The analysis included identifying the language-learning strategies that participants have used, the technologies they have used, and how they used VocabNomad. Participant attitudes towards VocabNomad were also identified.

6.2 Participants

Eight ELLs participated. FormEv1, FormEv3, FormEv5, FormEv7, and FormEv8 were in a post-secondary academic English training program. FormEv2, FormEv4, and FormEv6 were visiting
post-graduate students. Participant mean age was 23.1 years ($SD = 2.0$). One group of participants had been in Canada for just over a month (FormEv2, FormEv4, FormEv6), and the others had been in Canada for approximately four months. Participants received 10 dollars in exchange for their participation. Participants generally rated their receptive language skills (listening and reading) as being better than their language production skills (speaking). Additional participant demographics can be seen in Table 20.

6.3 Results

Participants demonstrated a variety of learning strategies. These strategies included their use of general and language-learning focused technologies. Participant interview data highlights how VocabNomad can be integrated into the learning strategies that ELLs already employ. This data also shows how the MALL approach that is operationalized through VocabNomad can be used to provide ELLs with additional learning opportunities.

6.3.1 Language-Learning Strategies

All but 1 participant (FormEv5) used English conversation to develop his or her language skills and thus demonstrated the use of languaging to support their learning. This included making friends with L1 speakers of English. FormEv2, FormEv3, FormEv4, and FormEv7 also noted that they try to think in English when speaking in English. FormEv6 did not take English classes when she was in school but was one of the students (FormEv2, FormEv4, FormEv6, and FormEv8) who sought private tutoring. All participants used some form of text-based studying to

---

**Table 20 Participant demographics – VocabNomad validation study.**

<table>
<thead>
<tr>
<th>ELL</th>
<th>Age</th>
<th>Sex</th>
<th>Mother Tongue</th>
<th>Language Spoken at Home</th>
<th>English Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$S^<em>$ $L^</em>$ $R^*$</td>
</tr>
<tr>
<td>FormEv1</td>
<td>22</td>
<td>F</td>
<td>Portuguese</td>
<td>English</td>
<td>5   9   5</td>
</tr>
<tr>
<td>FormEv2</td>
<td>27</td>
<td>F</td>
<td>Chinese (M)</td>
<td>Chinese (M)</td>
<td>6   8   7</td>
</tr>
<tr>
<td>FormEv3</td>
<td>23</td>
<td>M</td>
<td>Portuguese</td>
<td>Portuguese</td>
<td>5   8   9</td>
</tr>
<tr>
<td>FormEv4</td>
<td>24</td>
<td>F</td>
<td>Chinese (M)</td>
<td>Chinese (M)</td>
<td>4   6   5</td>
</tr>
<tr>
<td>FormEv5</td>
<td>21</td>
<td>M</td>
<td>Portuguese</td>
<td>Portuguese</td>
<td>7   8   8</td>
</tr>
<tr>
<td>FormEv6</td>
<td>24</td>
<td>F</td>
<td>Chinese (M)</td>
<td>Chinese (M)</td>
<td>6   6   7</td>
</tr>
<tr>
<td>FormEv7</td>
<td>21</td>
<td>F</td>
<td>Portuguese</td>
<td>Portuguese</td>
<td>6   8   8</td>
</tr>
<tr>
<td>FormEv8</td>
<td>23</td>
<td>M</td>
<td>Portuguese</td>
<td>Portuguese</td>
<td>7   8   8</td>
</tr>
</tbody>
</table>

$^*$ Self-reported, 1 = None, 10 = Perfect

M – Mandarin, S – Speaking, L – Listening Comprehension, R – Reading

---
Table 21 Participant technology use to support their English – VocabNomad validation study.

<table>
<thead>
<tr>
<th>ELL</th>
<th>TV</th>
<th>Radio</th>
<th>Music</th>
<th>Movies</th>
<th>Computer</th>
<th>Mobile Phone</th>
<th>Google Search</th>
<th>Tapes / CD</th>
<th>CALL / MALL</th>
<th>L1-L2</th>
<th>English</th>
<th>Electronic</th>
</tr>
</thead>
<tbody>
<tr>
<td>FormEv1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FormEv2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FormEv3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FormEv4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FormEv5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FormEv6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FormEv7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FormEv8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

improve their English. In some cases, this meant looking words up in the dictionary but it also involved reviewing grammar and practicing grammar rules. FormEv1, FormEv2, FormEv3, and FormEv7 would also watch TED talks to try and improve their English comprehension.

6.3.2 Technology Use

Participants had used a variety of technologies (Table 21) some of which were the same as those used in the proof-of-concept study (Chapter 3). It is worth noting that none of these participants listened to the radio in English or had used language-learning tapes or CDs. They had used MALL applications, primarily to support the development of their English pronunciation accuracy, which is a key difference between the participants in the MyVoice study and those in this study. Participants in this study used a greater variety of dictionaries or translation tools: this included the occasional use of Google searches to support their spelling or other vocabulary needs that they had noticed.

6.3.3 VocabNomad Use

All participants used VocabNomad to support their vocabulary acquisition by studying vocabulary entries, listening to vocabulary items and their associated sentences or definitions,
using the sentences to model appropriate word usage, or using the search function to learn how to spell words (see Figure 26). Participants reported that they found the images useful because the images helped them to connect the idea of the word with the picture of something that was already familiar to them, and FormEv5 reported that he had a preference for studying the words that were associated with images. FormEv1 and FormEv8 went beyond using the tool as a study aid: they also used it as a cognitive support for completing their homework.

Participants (FormEv1, FormEv5, FormEv7, and FormEv8) liked that they could choose when to listen to the expert pronunciation models that were provided through the text-to-speech engine. FormEv1, FormEv5, FormEv6, and FormEv8 also compared the recordings that they had made to the text-to-speech-generated pronunciation model.

All participants used the application in private locations and a subset of them (FormEv2, FormEv4, FormEv6, and FormEv8) used it in public settings that included a pub, the gym, a
research lab, and a grocery store. VocabNomad was primarily used to support vocabulary learning activities when ELLs used it in public. FormEv3, FormEv4, and FormEv6 used VocabNomad to support their communication, which enabled FormEv4 to get help finding a type of pumpkin when grocery shopping.

Participants explored learning content through the mobile and web-based interfaces. On the mobile client, they viewed the recommended synonyms or used the search function to discover new learning content or as FormEv8 expressed “you can search for words based on the topic”.

Participants found that the default content that was provided with their accounts was limited, which is why they chose to make requests for additional vocabulary. These requests supplied learners with new collections of words that had been generated using the mechanisms described in Chapter 4 and Chapter 5. However, participants reported that the algorithms took too long to return results. There was also some concern about noise in the data sets: FormEv8 noticed a word (prenatal) that seemed out of place in the collection of vocabulary that was tagged with gym.

Participants expressed a desire for a more collaborative approach to building learning materials and were willing to contribute to the development of the existing learning materials. FormEv3, FormEv4, FormEv6, FormEv7, and FormEv8 reported that they liked to add tags to vocabulary independent of whether they had added a new word or were building on vocabulary that was already in the system. FormEv3 limited his addition of tags to existing words and categories, whereas FormEv6 also added new categories of vocabulary to the system. The other participants who edited or added tags to vocabulary were adding new tags to existing vocabulary.

6.4 Discussion

Participants seemed content with the support that VocabNomad provided. They were able to use it to advance many of their language learning and some of their communication needs, with no clear indication of how their communication needs could be better supported.

Participants disliked how long it took the on-demand materials to be returned to their devices. This delay is largely due to network latency issues. Algorithm performance and a lack of previously collected data also contributed to the slowness that participants observed. However,
part of their displeasure is the result of a perception that participants had. This was revealed through a discussion with FormEv6 where she indicated that she thought it took too long for the system to find words that were on the device. Her response, when the interviewer told her that the system was going to the Internet and finding the words, indicates that changing the user interface to clarify where the words are coming from when the user performs a search would reduce user displeasure with the amount of time that is required for the system to provide these materials.

The interviews indicated that the addition of content sharing between users is desired by ELLs but that participants did not understand the mechanisms underlying the current implementation of vocabulary sharing. Participants did not understand the requirements that must be met for vocabulary to be shared. They expected anything that they shared to be pushed to all users. Participant misunderstanding that sharing is based on a push rather than a pull model, where learners must explicitly request the shared materials, may come from their interactions with social media, such as twitter and Facebook, where shared items are broadcast. In addition to the pull-based model, certain criteria needed to be met before vocabulary could be imported by others: vocabulary items needed to be shared, to have a tag associated with them, and at least some of the items that were associated with that tag were required to have an image. Additional information about the sharing and importing of vocabulary entries can be seen in 5.3.1 Domain Model.

6.5 Conclusion

Following the integration of content recommendation for emergent user needs and the validation of the system’s user interface, a deployment study was performed (N = 8). This study revealed the perceived usefulness of the VocabNomad approach for supporting vocabulary acquisition and communication.

VocabNomad’s design appears to have increased the flexibility with which users can organize and navigate support materials. It also emphasized VocabNomad’s ability to verbalize and record language. It seems that this may have helped convey that this approach to providing MALL support could be used to scaffold ELL communication since more participants used the application to support their communication during this study. However, it is possible that these
participants were more open to using technological communication supports than those from the deployment study that had used MyVoice.

Participant interviews indicated that ELLs appreciated the integration of on-demand vocabulary support but that it needs improving, especially with respect to its performance. This limitation may be partly overcome with a slight change to the user interface so that it better manages user expectations. Continued system use by ELLs will also result in the construction of vocabulary collections for different topics and contexts which will reduce the delay that participants experience when requesting support materials for emergent needs. In the future, it may be possible to incorporate predictive modeling of user content needs based on the previous support requests of other users.

The recurring request by participants to allow for the sharing of learning content between users indicated that VocabNomad’s sharing process needs increased transparency. Rather than redesigning the user interface, which would require additional validation, it was decided that we would improve user training. Since the ability to share learning content between users is not a focus of the research contribution for this system, the user interface for this feature can be improved when the collaborative processes of learners become a central focus of the work.

The only change that was made following this study which would be visible to users was the incorporation of fast-scrolling to the vocabulary list in the android client. A bug that was discovered in the logging of user actions within the system was fixed and several improvements were made to the android client’s performance. These improvements included the elimination of memory leaks and the addition of asynchronous tasks to move processes into the background, which improved system responsiveness.

The effort that was invested in designing and validating VocabNomad seems to have helped ensure a usable system that meets at least some of the vocabulary learning and communication support needs of ELLs (RQ1). User-satisfaction with the tool’s ability to support their language learning activities and communication indicated that it was ready for further evaluation. The above formative evaluation opened the door for conducting studies that explore the relationship between ELL use of an adaptive MALL approach, learner vocabulary knowledge (Chapter 7 and Chapter 8), learner affect (Chapter 8), and learner communicative success (Chapter 8) (RQ3).
We will begin with the exploration of ELL use of this approach in a formal learning environment, where we will focus on the development of ELL vocabulary knowledge.
Chapter 7 - Vocabulary Study in an English as a Foreign Language Class

While individual features were evaluated for their ability to provide specific types of support (Chapter 4) and VocabNomad was validated for its usability (Chapter 4), these studies provided no evidence of VocabNomad’s ability to support learning. VocabNomad’s formative evaluation showed that ELLs could use it to support learning activities that are linked with vocabulary acquisition, such as studying (Chapter 6). However, none of the prior evaluations tracked participant vocabulary knowledge so we have no evidence of how application usage and student vocabulary knowledge are related (RQ3). Moreover, the evaluation of VocabNomad as a whole has primarily been performed in informal educational settings; these evaluations involved everyday uses of the system rather than its integration into a classroom setting.

While this has helped answer RQ2 by informing our understanding of how ELLs integrate this MALL approach into their communication and language learning processes in informal language-learning contexts, it does not inform our understanding of how ELLs might use the application to meet their needs in formal educational contexts. I, therefore, explored the integration of VocabNomad into an English as a foreign language classroom with a focus on supporting existing course activities.

The integration of MALL tools into classroom settings has been limited. Previous work has shown that MALL tools can be effectively integrated into classroom-like settings when the activities performed have been a central focus of the curriculum (Jain et al., 2011; Wong & Looi, 2010). In many cases, students were given mandatory tasks, such as identifying situations where idioms could be used (Wong & Looi, 2010) and were required to perform group work while using the MALL tool (Jain et al., 2011; Wong & Looi, 2010). Other classroom integration projects have focused on the use of MALL as a supplementary activity to support student exposure to the language and allow student travel to be incorporated into a course (Demouy & Kukulska-Hulme, 2010; Kukulska-Hulme & Bull, 2009).

The use of MALL to support regular classroom activities is not well understood with teachers having been given little guidance on how to effectively integrate MALL into their curriculum
(Beatty, 2013). As a result, the integration of MALL into a classroom environment poses risks to student learning. Most MALL studies last for periods of time that are inconsistent with the duration of a course or its curriculum: more than 70 percent of MALL evaluations have lasted less than a school term (Burston, 2014b). This study explores the integration of VocabNomad into a classroom over the course of an entire term (RQ2). A desire to prevent the integration from negatively affecting student learning led to a more collaborative approach and a methodology that is consistent with action research was followed.

This approach was chosen because it is a disciplined inquiry process that involves stakeholders in the research process rather than limiting their role to that of participant (Sagor, 2000). This does not mean that all participant-researchers are given the same amount of power and control, rather there is a spectrum of power and control with different stakeholders negotiating different positions on that spectrum (Smith, 2012). For this study, the teacher partners and I shared a similar level of control. The school and its administrators exercised slightly less control but could have terminated the study. The school was also given the authority to review manuscripts prior to publication. Students were given the least control over the study even though their actions affected it the most.

Action research traditions often more closely align themselves with qualitative inquiry activities (Smith, 2012), such as traditional program evaluations, but they can be paired with quantitative methods of inquiry. Rather than strictly relying on either of these research traditions, a mixed-methods approach was employed because it allowed for reliable quantitative measures to be taken without completely losing the voice of researcher-participants. The methods for evaluating VocabNomad’s integration into the classroom (RQ1 and RQ2) were more qualitative with details about the study context and the interpretation of qualitative data undergoing member checking. The methods used to evaluate the relationship between VocabNomad usage and student language abilities (RQ1 and RQ3), including vocabulary knowledge, were quantitative and did not undergo member-checking.

A comparative study was not planned because we do not yet know the relationship between VocabNomad usage and ELL vocabulary knowledge. Moreover, few MALL tools have been shown to result in vocabulary growth (Burston, 2014b) and even fewer adapt their content to
learners while still allowing the learner to exercise a great deal of control. This makes it difficult to find an appropriate comparison system and indicates that a comparison study might be inappropriate at this time (Paramythis et al., 2010). As a result, the intent of this study was not to prove that one tool outperformed another but to study how VocabNomad affected the classroom environment and student learning.

This study focuses on two sub-questions of the primary research questions:

**RQ4.** To what extent can we effectively integrate VocabNomad into an English as a foreign language classroom, when using it as a supplement to other course materials and activities?

**RQ5.** How does VocabNomad use relate to student vocabulary knowledge?

Answering RQ4 contributes towards the larger question about how users integrate these types of tools into their language-learning activities (RQ2). RQ5 helps answer the larger research questions of the relationship between system usage and student learning (RQ3) as well as the extent to which the tool can support vocabulary learning (RQ1).

From these questions the following hypothesis was tested:

**H1:** ELL receptive vocabulary will increase following the introduction of an adaptive mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary use.

Test results indicate that student vocabulary knowledge increased during the period when students used the developed tool the most, and the collected data show us that VocabNomad was successfully integrated into an English as a foreign language classroom. This indicates that similar tools could be integrated into classroom settings provided the appropriate expertise for overcoming social, cultural, and technological barriers to using the tool is available.

I first explain the manner in which the study was conducted. This includes the introduction of the tool (i.e., VocabNomad) and the researcher (i.e., myself) to the research site as well as a description of the data collection procedures. The research site is described and the study results are detailed; this includes the challenges, successes, and opportunities that presented themselves.
The results are followed by a discussion of the recommendations and reports that were provided to individual participants and the school.

7.1 Methodology

A mixed-methods design that is consistent with an action-research methodology was used because it allows for greater flexibility. This flexibility comes from a reflective process that allows both researchers and participants to contribute to the development and adaptation of the study methods. This approach has facilitated exploratory research projects that were conducted in educational settings where power differentials among researchers, teachers, and students may negatively affect the project (Avison et al., 2001). The power differential between the students and the teacher-researchers, including myself, was largely preserved as a result of Japanese cultural norms with respect to the roles that people are expected to play (De Mente, 2009). Students were allowed to take the devices home with them, which would enable students to exercise greater control over the devices. The decision to let students take the devices home also held the potential to enable the completion of additional learning activities.

The study methods and goals were first presented to the school via a document that was sent to a teacher (Sensei_1) at the school who was a friend of mine (Sensei_3). It was then passed on to his department head and once she had approved of my visiting the school, the information was passed on to her superiors for approval. The high-level study goals were explained to the students when they were given the devices on the first day of class.

In keeping with the cultural norms of Japan, small gifts of ice wine candies were brought and given to the teachers in the English department and maple candies were brought to give to the students on each of the testing days. Gifts of ice wine and ice wine tea were given to my host teacher (Sensei_1) and his wife who coordinated my local living arrangements and picked me up from the airport.

To honour the action-research methodology, Sensei_1 and the school chose the course, English Intensive 2 (El2), in which the deployment would occur; the desire of the other course teacher (Sensei_2) to find new ways to develop student vocabulary knowledge contributed to this decision. I began to coordinate with Sensei_1 and Sensei_2 in December, approximately four
months before the class was scheduled to begin. We worked together to ensure that the introduction of VocabNomad would integrate into the current course syllabus without having to change the primary focus of the course. Teachers participated in the design of the learning activities for which VocabNomad was to be used as well as the data collection and interpretation. In general, a consensus-based approach to decision making was used.

As several researchers have found (Avison et al., 1999; Dubé & Paré, 2003; Lau, 1997), this collaborative approach along with the teacher-researcher’s continued involvement in the study’s evolution can result in participants wanting to see positive outcomes as a proxy for their successful participation. To account for this potential bias and to measure changes in student vocabulary knowledge, various tests were administered and the data that resulted from the automated logging of user actions within VocabNomad were analyzed. Classroom observation was performed by all three teachers, and all testing was coordinated to ensure that student participation would not affect their other courses or co-curricular activities.

The original study design involved the removal of the application at the study midpoint. However, limitations with respect to the WiFi connectivity that were discovered during class resulted in our deciding to let the students keep the application for the entire term. We then tried to recruit students from other courses to serve as a control group. These students participated in the administration of the tests for vocabulary and morphological knowledge in return for receiving a detailed report about their abilities. These students did not use VocabNomad. The recruitment of the control group was performed by their normal classroom teachers, who announced the special test times. It was made clear that student participation was voluntary.

7.1.1 Researcher Introduction and Integration

I am referred to as Sensei_3 throughout the remainder of this document. I was introduced to the school by Sensei_1 since we had a prior relationship. My curriculum vitae and a package that included information about the system, my approach to research, and the proposed research plan were given to the English department head and school principal by Sensei_1. After the school agreed to participate in the research project, a second teaching partner, Sensei_2, was selected and introduced via email.
Initial planning of the tool’s integration into the selected course began approximately four months prior to the start of the study and took place over email. I arrived at the school approximately one week before courses began and continued to work with Sensei_1 and Sensei_2 to prepare for the course. All three of us participated in the preparation of course materials with Sensei_2 preparing the majority of lesson plans and worksheets.

Upon arrival, Sensei_3 was introduced to all of the teachers in the English department. Sensei_3 was also introduced to the teachers in the school at a meeting prior to the school’s commencement ceremony, where she was introduced to the students.

Sensei_3 worked at a desk in the English department during regular school hours and attended the majority of the weekly student-led assemblies. Sensei_3 also attended the weekly school religious services and participated in the English speaking society. This involved interacting with students and teachers at two lunch and three after school meetings each week. Lunch meetings lasted approximately 45 minutes and the after school meetings lasted between 60 and 90 minutes.

7.1.2 Tool Introduction, Explanation, and Training

VocabNomad was introduced during the first class of the term. An explanation of the research, the application, and its development was given in English by Sensei_2 and Sensei_3. This included a statement by Sensei_2 that Sensei_3 had built the application.

The devices and a user manual were distributed, and the application features were demonstrated. Sensei_1 was away during this class. So, he re-explained my presence and the purpose of the study, using Japanese, during the next class. We also demonstrated more features during the second class.

7.1.3 Consent and Remuneration

Even though the school had consented to participating, student consent was voluntary. Consent was explained in English and Japanese. Students could choose to explicitly communicate that they did not wish to participate. Due to cultural expectations this was unlikely to happen, but students could refuse to use the application or to answer test questions. The control group that
was recruited mid-way through the study had the option of not coming to subsequent testing sessions.

Students received no monetary remuneration for their participation. Candies were distributed on test days and all students, even those who withdrew, were given both a report detailing their results and a personalized workbook that was filled with activities to help develop their English-language skills.

In keeping with an action research approach, the school also received a report on the students’ abilities with suggestions for where improvements could be made (see 7.9 Stakeholder Reports for more information).

7.1.4 Data Collection

Four types of data collection were performed: 1) user observation, 2) demographic information, 3) learner self-reports of language proficiency, and 4) measures of learner language knowledge. Additional cognitive and language measures had been planned but were cut, in consultation with Sensei_2 and Sensei_1, for logistical reasons. These measures included the adaptive vocabulary testing that is part of VocabNomad, measures of phonological awareness and processing, and measures of learner’s working memory and recall. The primary constraint that limited the use of these measures was a time limitation that resulted from students taking longer than expected to follow the instructions that they were given in class.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>11 April 2014</td>
<td>Treatment</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 April 2014</td>
<td>Treatment</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>13 May 2014</td>
<td>Control</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>20 May 2014</td>
<td>Treatment</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 June 2014</td>
<td>Control</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 June 2014</td>
<td>Treatment</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04 July 2014</td>
<td>Treatment</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
</tbody>
</table>

Black circles indicate when a measure was taken for each of the student groups.
Demog. – demographics, Profic. – proficiency, Orth. – orthographic, Morph. – morphological
The timing of data collection for each of the measures that were not based on observations is detailed in Table 22. The measures listed in this table are defined throughout this section. The demographics forms and measures of morphological knowledge were administered once. The measures for vocabulary knowledge and self-reports of language proficiency were collected at three time points.

7.1.4.1 User Observation

Two types of user observation were performed. The first was the automated logging of every action within the application, regardless of the platform that was being used. These logs were transferred between the server and android client at regular intervals to ensure that data loss was kept to a minimum (see 5.3.3.1 Data for details about the application logging). Classroom and other learner observation was performed by the participating teachers and myself. There were typically three observers in the classroom; the first class was the only class where two observers were present. Reports from other teachers who noticed students using the application or who overheard them talking about it were included within the data.

Figure 27 shows the facilities that we were using and an image of how each desk appeared when students arrived to the first class. The class was conducted in the media room which had 6 rows of 8 tables with laptops that were running the Japanese version of Windows 7. All laptops had wired Internet connections.

7.1.4.2 Demographics Information

The demographics form collected information about learners’ mother tongue and exposure to English. This included family member knowledge and use of English. The form also collected information about the time that students had spent in English-language environments, the English-language media that they use, and the other learning activities that they had performed or were currently performing. Students were asked to rate their proficiency in English and any other languages that they know. A copy of the demographics form can be seen in Appendix B.

Sensei_2 helped answer student questions about the demographics form when we were filling it out during the first class. These were reviewed and inconsistencies or other data collection problems, such as incomplete information or unclear birthdates were identified. The Japanese
teacher, Sensei_1, helped get clarification from students about some of their responses during subsequent courses. The English language self-assessments were the first to be clarified. Other data were clarified or confirmed over the course of the school term; there was no specific order to the clarification of the other types of data.

7.1.4.3 Learner Self-reports of Language Proficiency

After reviewing the students’ responses to the demographics forms, we decided to collect self-assessments of their Japanese-language abilities when we administered the first set of tests. We also decided to re-administer the English and Japanese self-assessments with the mid-point and final vocabulary tests (i.e., PPVT-4).
7.1.4.4 Assessments of Language Knowledge

The vocabulary assessments (PPVT-4 and a teacher-created final exam) were used to track participant vocabulary knowledge, whereas the other tasks and tests were used to measure different variables that can affect vocabulary acquisition. The primary measure of language knowledge, the PPVT-4, was administered at the beginning, in the middle, and at the end of the study (see Table 22). Other variables were measured once, with the measures of morphological knowledge being administered alongside the initial PPVT-4 test and the teacher-created final exam following the final administration of the PPVT-4.

7.1.4.4.1 Peabody Picture Vocabulary Test 4 (PPVT-4)

A group-based administration of the fourth version of the Peabody Picture Vocabulary Test (PPVT-4) was used to track changes in learner vocabulary knowledge. The administration method that was used was similar to that used by Ramirez et al. (2009). Every third item from the test was administered (e.g., 1, 4, ... 226; 2, 5, ... 227; and 3, 6, ... 228) to maintain the progression in difficulty from item to item. Students were shown the images using the projector at the front of the class and the stimulus word was read two times before advancing to the next test item. Figure 28 provides an example of the general visual stimulus template.

This administration method allowed for the use of a single test at multiple points while reducing the potential for a false positive as the result of student familiarity with test items. However, it meant that the basal and ceiling values could not be determined so raw scores were used.

Instructions were given in English and Japanese, and the training tasks were modeled by Sensei_1 during the initial test. He gave the Japanese version of the instructions. Subsequent administrations were performed by myself with Sensei_1 and Sensei_2 observing the students and providing support.

7.1.4.4.2 Teacher-created Final Exam

This portion of the term-end exam was created by one of the classroom teachers (Sensei_1). All of the vocabulary items that were tested were taken from those provided in VocabNomad. Students were asked to complete sentences by filling in the blanks using items from a word bank, to match words with their definitions, to produce the word that matched a definition, and to
Figure 28 An example of how a PPVT-4 stimulus page might look for the word sun. This example is fabricated to indicate the nature of the test while preserving its integrity.

reorganize a list of words or phrases into a comprehensible sentence. See Appendix I for a copy of the full exam. Only student results from the vocabulary portion of the exam are reported.

7.1.4.4.3 Morpho-Orthographic Choice Task

Students completed this task independently. It is a receptive measure of their knowledge of English spelling rules and how words change when suffixes are added. It asks students to recognize and circle the correctly spelled word from a pair (e.g., complition and completion). The task items can be seen in Appendix I.

Instructions were given in English and Japanese. Some students confirmed the task instructions using English.

7.1.4.4 Carlisle’s Morphological Structure Test

This task asks participants to modify a word so that it completes a sentence (Carlisle, 2000). For example, a student would be given the word farm and the sentence “My uncle is a ________”. 
The student is expected to provide farmer as an answer. This example is one of the training items from the task. The full task can be seen in Appendix I.

The administration of this task was modified so that it was appropriate for use in a group setting. Rather than reading the sentences to participants and having them respond verbally, participants were given a copy of the test form and asked to write the appropriate transformation of the target word. The scoring method was modified so that correct responses earned 2 points and responses that were correct but contained spelling errors earned 1 point.

Once again the instructions were given in English and Japanese, and Sensei_1 modeled how to complete the training tasks.

7.2 Research Site

The deployment study was conducted at a private all-boys Christian senior high school. The school is affiliated with a university, and unlike many other Japanese high school students, the students at this school do not need to worry about their ability to gain admission to university or the need to write the high-stakes entrance exams that are typically required to gain admission to university in Japan. Students who wish to attend a different university still write these exams but those who decide to attend the university that is affiliated with their school are not required to do so. This alleviates the need to follow a set curriculum and gave us the ability to explore the integration of VocabNomad into a classroom setting since no high-stakes exams depended on student learning outcomes.

The course that the school selected for the deployment was English Intensive 2 (EI2). Students in this course receive a minimum of 4.5 hours of English-language instruction per week. It is an elective course that students can take in their final year. Its curriculum focuses on developing student presentation skills: students learn how to prepare PowerPoint slides and give group presentations. This offering focused on famous foreigners since Sensei_1 felt that this would appeal to student interests; he knew many of them from his homeroom or from the school baseball team that he managed.

Students at this school must take Japanese, English, and Math tests at the beginning of each school year (the beginning of April) and are placed into streams based on the results of these
tests. The top performing students are placed into the science stream, those whose scores place them in the middle of the cohort are assigned to the English stream, and those whose scores place them in the bottom of their cohort are placed in the social sciences stream. The science stream students are segregated whereas the students from the other two streams are integrated during their English courses. In addition to this, students take university transition courses at the affiliated university during their third and final term of high school.

Because it was decided that VocabNomad would only be used in one classroom, other courses were observed to better understand the culture of the school and the manner in which it approaches English language instruction. This gave me a comparison point for the course where we were doing the deployment.

Sensei_2 and his Japanese teaching partner (Sensei_7) allowed me to observe some of their other courses, which were at the second-year level. Sensei_2 and Sensei_7 selected two courses that they had identified because they were representative of how the non-grammar courses were run. So, I sat in on the same lesson while it was being delivered to two student cohorts: the first was the science stream cohort that the teachers considered the rowdier of the two and the other was judged to be a typical mixed-stream cohort.

Sensei_1’s teaching partner (Sensei_5) consented to my observing some of their third-year courses. This provided a comparison point for how the instructional staff and student cohort may have affected student behaviour in EI2 since the EI2 students were mixed in with the other third year students for the courses that I observed. In this case, I watched different lessons being delivered to two different subgroups of the third-year student population.

7.2.1 Course Organization: EI2

The course in which the deployment took place was held twice a week: once on Monday mornings in second period (9:25-10:10) and once on Friday afternoons, immediately after lunch (1:15-2:05). Each course started with a group greeting and bow. Student homework requirements and other administrative tasks were performed when necessary and an overview of the day’s plan was provided. The lesson plan typically involved some form of worksheet or group work activity. Lessons ended with a farewell message and bow once the end of period bell had rung.
The greeting and bowing ritual was initiated by the Japanese speaking teacher (Sensei_1); he would prompt the student leader (Nihon_S16) who then led the rest of the group in this ritual. While many independent work opportunities were given to students, the greater majority of classroom practices were teacher centered.

7.2.2 Course Organization: Other Courses

Like EI2, the other second and third year course activities were primarily teacher centered even though they involved a considerable amount of student activity. Individual classes started and ended with the same student-led greeting and bowing ritual. Vocabulary review and some form of interactive activity were conducted in each class. Group work, typically performed in pairs, was also employed. These classes seemed to move at a far faster pace, with multiple activity changes, than the more project-based EI2 classes.

All of the activities that were performed by both sets of teaching partners were highly structured. Students, in all courses, were given the opportunity to participate by interacting with classmates or by answering teacher questions. The courses that I observed were conducted using a combination of English and Japanese, with the majority of teacher speech being conducted in English.

Additional information about the course organization for the observed second and third year courses can be seen in 7.2.4 Learning Activities: Other Courses.

7.2.3 Learning Activities: EI2

The term started out with smaller activities that were grounded in teacher-prepared worksheets. These activities were intended to prepare students for their term project which involved researching, preparing, and delivering a group presentation on a famous foreigner.

Vocabulary was distributed to students via VocabNomad. Initial vocabulary sets were accompanied by worksheets where students were expected to use VocabNomad to determine the equivalent Japanese word and then check their responses using a dictionary. The vocabulary chosen for these worksheets was meant to support students’ ability to work in PowerPoint or perform other group and computer-based activities using English. It was also meant to support
their understanding of the content from Sensei_1’s model presentation. Some of the worksheets that were used during the term can be seen in Appendix J.

Prior to giving their presentation, students were required to prepare vocabulary lists where each word was accompanied by a sentence using that word. The students were instructed to select vocabulary from their presentation that they thought their classmates might not know. These vocabulary were initially distributed via VocabNomad, and a paper copy of a subset of the vocabulary that they had been given over the term was given to students prior to their final exam.

Students were reminded that the vocabulary had been put into VocabNomad for their review during the class immediately prior to the presentations for which they would need that vocabulary.

7.2.4 Learning Activities: Other Courses

The other courses used a variety of activities that centered on student readers, workbooks, and teacher-prepared handouts. Shadowing, the practice of reading aloud with an audio recording of a text (Marslen-Wilson, 1985; Shockley, Sabadini, & Fowler, 2004), was encouraged and modeled by some teachers.

HyperListening (Kiriharashoten editing section, 2012) was used throughout the school to help students develop their listening skills. This involved students listening to a pre-recorded text and answering questions.

The portion of the class that was dedicated to vocabulary review for both courses involved a repeat-after-me activity where students would repeat the words after the Japanese-born teacher said them. Additional vocabulary review involved the translation of English words into Japanese at various points throughout the class. The third-year teachers (Sensei_1 and Sensei_5) would also provide examples of words or model their use so that students could figure out what the equivalent Japanese word was. In some cases, Sensei_1 would model word decomposition to help students determine the meaning of a word.
7.3 Participant Demographics

We first present information about teacher demographics. Demographic information is only provided for the teachers who contributed to the collection and interpretation of data. This is followed by information about the control and treatment groups.

7.3.1 EI2 Teachers

Sensei_1 was Japanese born and has been teaching English at this school for eight years. He had completed Bachelor and Master of Arts degrees at Japanese institutions. His Master of Education was completed at a North American institution. He had also earned a Teaching English to Speakers of Other Languages (TESOL) certificate. He is fluent in both English and Japanese and has spent two and a half years in English-language environments.

Sensei_2 was born in an English-language dominant country and has been teaching English in Japan for ten years. His teaching experience spans many different educational contexts within Japan and this was his second year teaching at this school. He has completed a Bachelor of Arts and is currently pursuing a Master of Arts. His mother tongue is English and he rates himself as an intermediate speaker of Japanese.

I (Sensei_3) was born in an English-language dominant country. I had previously taught conversational English as a second language for a year. This was my first time teaching in a Japanese context. I have completed bachelor and master’s degrees and was pursuing my PhD. I am fluent in English but have little knowledge of Japanese.

7.3.2 Other Teachers

Sensei_4 was European born and has been at this school for over 10 years. He has lived in Japan the majority of his life and speaks English and Japanese fluently. He had completed bachelor and master’s degrees in North America.

Sensei_5 speaks fluent Japanese and English. He was born in an English-language dominant country and has been teaching English in Japan for over 10 years. This was his second year teaching at this school and his teaching experience spans various educational contexts in Japan. He had completed a Bachelor of Arts degree and Teaching English as a Second Language
(TESL) certification. He is working towards a Master of Education degree and is one of Sensei_1’s teaching partners.

Sensei_6 was European born and has been teaching English at this school for five years. She has completed Bachelor and Masters of Arts degrees and is pursuing her doctorate. She is fluent in Japanese and English. She also teaches Japanese as a foreign language.

Sensei_7 was Japanese born and has been teaching English at this school for over ten years. He has a Bachelor of Arts degree from a Japanese institution and a graduate diploma from an English-language institution. He has spent two years in a country where English is the dominant language. He is fluent in Japanese and English. He is one of Sensei_2’s teaching partners.

7.3.3 Students

Treatment group students are identified using a moniker that begins with Nihon_S and is followed by a number (e.g., Nihon_S01). Instead of an S, student monikers for those who belong to the control group use a C (e.g., Nihon_C01). All students identified Japanese as their mother tongue and as the language that they spoke at home. Detailed information about student exposure to the English language outside of their school environment can be seen in Appendix K.

All 47 treatment-group students, aged 17.46 years (SD = 0.40), were enrolled in a third year course that is called English Intensive 2 (EI2). While only 2 treatment-group students had used English with a parent (Nihon_S16 and Nihon_S37), 14 of the students had at least one parent who speaks English (8 mothers, 9 fathers); Nihon_S37 used English when travelling with his parents, and Nihon_S16 estimated that he uses English with his father about once a week. Only 1 student, Nihon_S45, had someone else at home who knew English, but he did not use English with his sister.

Four students (Nihon_S08, Nihon_S21, Nihon_S33, and Nihon_S46) had lived in English language environments for more than a year (see Appendix K): English was the only language that Nihon_S21 shared with people outside of his family when living in Singapore. Two students from the control group (Nihon_C05 and Nihon_C11) and 13 of the treatment group students had or were performing extra-curricular English-language learning activities.
Twelve of the 13 students from the control group were in their third year. The last member of this group (Nihon_C13) was a second-year student who was enrolled in English Intensive 1, which indicates that he is one of the top English-language performers in his year. The average age of control-group students was 17.52 (SD = 0.49) years. The father of Nihon_C07 knew English. However, students only reported using English with their parents occasionally. Seven of these students (Nihon_C01, Nihon_C03, Nihon_C05, Nihon_C07, Nihon_C11, Nihon_C12, and Nihon_C13) completed the study. All summary statistics and tests only include these 7 students.

All of the English-language environments that students had visited or lived in were members of the Pacific Rim, with American destinations being the most common: 14 of the 21 students had visited or lived in at least one American city. The next most popular locations were Guam (6), India (4), and Australia (3).

Many of the students used English language media: 47 regularly listened to music in English and 37 watch television or have seen movies in English. Only 2 students (Nihon_S04 and Nihon_S13) read English language texts.

If students are categorized into three groups (high, medium, and low) based on their exposure to English, we see a moderate correlation ($r_s(45) = .536, p < .01$) between student exposure to English and the ability of at least one of their parents to speak English. While this might be expected, it is worth noting that student exposure to English did not include its use with their parents. Given the relationship between student exposure to English and their parents’ knowledge of English, parental English serves as a proxy for the family’s acceptance of the importance of learning English and less traditional, from a Japanese perspective, methods for learning it.

The above data highlight the variability in student backgrounds within the course where VocabNomad was deployed. These data also demonstrate that students from the control group had similar backgrounds to those in the treatment group.

7.4 Results

I first describe how students reacted to the use of smartphones and VocabNomad in their class. This includes how they used the application. Student learning gains are shown through their
performance on English-language assessments. Gains in their self-perceived English language abilities are detailed before the relationship among system usage, student language abilities, and student learning is described using the results of a discovery with models analysis.

7.4.1 Student Reactions to the Devices

Sensei_2 and Sensei_3 noted students’ initial excitement over the introduction of smart phones into their class. One student was completely shocked and made a considerable amount of noise when he saw the Samsung Galaxy Nexus on his desk. When asked why he reacted that way, he reported that it was because he “love[s] Google”.

Student excitement over the application and devices was demonstrated by their playing with the text-to-speech within the application during class even though the behaviour was, technically, off-task. They also showed it to the students and teacher (Sensei_4) in a neighbouring classroom during the break between courses.

Given these reactions and the students’ extensive use of their own smartphones outside of the school context, it is not surprising that students personalized their devices by adjusting several settings. The details of this customization can be seen in Appendix K.

The device language was set to Japanese by 19 students. The set language on the remaining 28 devices was English. However, some of the students changed the type of English that was specified. Only 2 of the Moto G devices had been handed out with American English set as the default language; all other devices had been set to use Canadian English. However, 1 device was returned with developer accented English, 3 specified United Kingdom English, and 17 specified American English.

Some students modified their keyboard and other input settings. The keyboard on 14 (29.8%) of the devices was set to the iWnn IME which displays the Hiragana character set, the Roman alphabet, and Arabic numerals to support both English and Japanese input; 4 students (8.5%) enabled Chinese pinyin support; 1 student (2.1%) enabled Hindi transliteration even though he does not know Hindi; and Nihon_S07 installed and enabled Japanese emoji while still using an English keyboard. The voice typing and spell checker features were left on by 45 students (95.7%); Nihon_S39 turned spell checking off and Nihon_S24 turned voice typing off.
Only Nihon_S40 appeared to have recorded a video, and 27 students (57.4%) took photos using the device camera. Nihon_S31 took the most photos (136). On average, students took 6.57 (SD = 20.83) photos.

Twenty-nine students changed the applications that were on their homepage:

- 3 saved VocabNomad to the homepage,
- 6 added the analog clock,
- 6 added the Google Play store,
- 5 moved the camera from the application bar onto the homepage, and
- 5 changed the camera in the application bar to the one that integrates with Picassa.

Students installed an average of 5.07 (SD = 9.62) applications, with at most 32 being installed by Nihon_S38. Google translate was only installed by Nihon_S43 and Nihon_S31. At least 8 students (Nihon_S01, Nihon_S07, Nihon_S10, Nihon_S13, Nihon_S21, Nihon_S31, Nihon_S38, and Nihon_S43) linked their gmail accounts to the devices and one student (Nihon_S19) appeared to have performed a factory reset on his device before reconfiguring it and returning it. Another student (Nihon_S26) used Picassa through the account that was associated with his mobile device, and 3 of the students who had linked their gmail accounts (Nihon_S07, Nihon_S13, and Nihon_S21) used those accounts to access Picassa.

Several students used their devices to perform Internet searches and visit web pages as is evidenced by nine students returning their devices with up to 5 browser tabs left open. The files that were left in the downloads folder of Nihon_S13 and Nihon_S20 provide further evidence of this. These activities were at times related to assigned learning tasks since some of the searches and at least one of the downloaded files were related to the presentation topics that had been selected by students (Figure 29).

7.4.2 Student Reactions to VocabNomad

Students showed surprised expressions when they were told that Sensei_3 had built the application that they would be using. They also demonstrated excitement over the use of technology in the English class but were concerned about how the application’s use might affect the course and their grades. This was demonstrated by one student approaching Sensei_2 at the
Figure 29 Examples of student on-task behavior: translating a word that Nihon_S25 needed for his presentation (centre), and a customized home screen where the background image is related to Nihon_S13’s presentation topic (left). The camera icon indicates that Nihon_S13 also customized his device by integrating his Gmail and Picasa accounts. An image that Nihon_S20 downloaded (right); it related to his presentation topic.

end of the class to discuss this issue. Some students also expressed concern or confusion over what they would be expected to do with it. At this point, we had not assigned specific tasks where they would use the application but had only described how it could be used and expressed that we would be using it. Another student, much to the surprise of Sensei_1, talked to me at the end of the first class and came to see me in the English department to get help with using the application to complete the worksheets. The identity of these students is unknown because this happened on the day where Sensei_1 was away and neither I nor Sensei_2 had learned student names at this point.

7.4.3 VocabNomad Usage

We first report on usage that was observed through the automated logging of participant activities within VocabNomad. We then report on activities that were observed by humans.

7.4.3.1 Logged Usage

The log files show an order of magnitude difference between the number of learner actions that were recorded in the first and second half of the study (see Table 23). Table 23 indicates that all students used the application at some point, even if their usage was minimal, since the minimum usage for the entire study across both phases was 84 events.
Table 23 Student application use (number of logged events) by study phase.

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>30</td>
<td>0</td>
<td>84</td>
</tr>
<tr>
<td>Q1</td>
<td>869</td>
<td>0</td>
<td>1096</td>
</tr>
<tr>
<td>Md</td>
<td>1630</td>
<td>45</td>
<td>1648</td>
</tr>
<tr>
<td>Q3</td>
<td>3082</td>
<td>230</td>
<td>3362</td>
</tr>
<tr>
<td>Max</td>
<td>9746</td>
<td>2482</td>
<td>9746</td>
</tr>
</tbody>
</table>

The Shapiro-Wilk test for phase 1 usage ($p < .001$) and phase 2 usage ($p < .001$) indicates that students’ daily event counts were not normally distributed, and a significant reduction in usage was observed between the first and second phases of the study ($Z = -5.87$, $p < .001$). The first phase is the period between the introduction of VocabNomad and the midpoint administration of the PPVT-4. The second phase is the period between the second and third administrations of the PPVT-4. Phase 2 is also when students were encouraged to use VocabNomad to review vocabulary but were not given explicit exercises that required application use.

Nihon_S41 used VocabNomad at the same low-level throughout the entire study. Nihon_S49, Nihon_S30, Nihon_S06, and Nihon_S01 reduced their application usage by 50-80% but continued to use the application at a level that was similar to that of Nihon_S41 during phase 2. Nihon_S34, Nihon_S35, and Nihon_S37 maintained moderate to high levels of usage during phase 2 even though 50-75% fewer events were logged. The other 39 students either reduced their usage to a level that indicates fewer than 5 minutes of usage per week or stopped their usage altogether during phase 2.

The statistics presented in Table 24 are based on the types of events that the system logs, which are defined in Chapter 5 (Table 17 and Table 18). These logs show that students used both English and Japanese to find vocabulary within the application. The logs also show that students predominantly used the application to study vocabulary (i.e., Viewing Vocabulary row in Table 24). Students seem to have appreciated being able to listen to pronunciation models as shown by their listening to the verbalization of sentences through the text-to-speech engine (i.e., Listening to a Sentence row in Table 24) or by listening to a recording that they had made (i.e., Recording Playback row in Table 24).
Table 24 Selected usage statistics by quartile.

<table>
<thead>
<tr>
<th>Event</th>
<th>Min</th>
<th>Q1</th>
<th>Mdn</th>
<th>Q2</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese Search</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>86</td>
</tr>
<tr>
<td>English Search</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>45</td>
</tr>
<tr>
<td>Viewing Vocabulary</td>
<td>0</td>
<td>0</td>
<td>805</td>
<td>1411</td>
<td>3058</td>
</tr>
<tr>
<td>Adding Vocabulary</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Editing Content</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>Deleting Content</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sharing Vocabulary</td>
<td>0</td>
<td>15</td>
<td>24</td>
<td>34</td>
<td>84</td>
</tr>
<tr>
<td>Unsharing Vocabulary</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Recording</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>Recording Playback</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>Listening to a Sentence</td>
<td>0</td>
<td>4</td>
<td>13</td>
<td>31</td>
<td>154</td>
</tr>
<tr>
<td>Requesting Vocabulary Just in Time</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>17</td>
</tr>
</tbody>
</table>

Students shared vocabulary items (i.e., Sharing Vocabulary row in Table 24), which is expected given the collaborative nature of many of the tasks in the EI2 classroom, but they did not directly add many new vocabulary entries (i.e., Adding Vocabulary row in Table 24). Few items were made private (i.e., Unsharing Vocabulary row in Table 24) and sharing actions were only done through the android client. In contrast, the deletion of tags or vocabulary items was done through both platforms (i.e., Deleting Content row in Table 24). Similarly the editing of learning content was performed through both clients (i.e., Editing Content row in Table 24). Some students performed L1 searches, which would trigger the entered item to be translated from Japanese to English. Only Nihon_S22 entered Japanese text into any part of a vocabulary item.

To better characterize student usage patterns, models of system usage were built using RapidMiner 5.3.015 (RapidMiner, 2014). K-means clustering was applied to the weekly usage statistics of individual students to identify usage patterns and group students based on those patterns. Model features consisted of each student’s unitized event counts for the week. This means that each student had a usage value between zero and one for each week of the study. K values of 2, 3, 4, 5, and 6 were attempted. In each case, both the maximum number of optimization steps and runs were set to 1000 and a random seed was used.

Visual inspection of the centroid plots and cluster assignments led to the rejection of the clusters that were created using 2, 5, or 6 for K since these clusters did not add meaningful information to
those that were formed when K was set to either 3 or 4. With K set to 3, a Davies-Bouldin Index (BDI) of 0.93 was observed and the following usage patterns were identified (Table 25)\(^9\):

- **Excited users (C_E_U):** these participants had high usage levels early in the term. Their usage quickly dropped and most learners in this group had very low usage during the second half of the term.
- **Just-in-time users (C_J_U):** these users had moderate levels of usage early on and returned to using the application when they felt it was necessary or course-related activities, such as exams or preparing vocabulary lists, encouraged MALL tool use.
- **Minimal users (C_M_U):** these participants had low initial usage levels. Students in this group either maintained their low usage levels or stopped using the application.

With K set to 4, the BDI decreases to .88 and another usage pattern emerges (Table 25):

- **Consistent users (C_C_U):** the participants had reasonably high initial usage which dropped more slowly than that of the excited users. They also used the tool regularly and reviewed vocabulary around the same time as their exams and their classmates’ presentations. They did not use the tool during the weeks when their essays were due.

The decrease in DBI indicates that the clusters that were identified when K = 4 are more cohesive and further apart than those that were identified when K = 3. The patterns that are visualized in Table 25 show the average number of events (on a log scale) during each study week for the students assigned to the identified cluster. The student assignment to clusters (Table 26) also shows increased refinement in the clusters that were identified when K was set to 3, and Table 27 shows that the change in centroid values is not as dramatic as might be expected if one were only looking at the cluster assignments: the centroids for weeks 2 and 3 are the only ones that showed substantive value changes.

\(^9\) The labels assigned to each cluster are based on my interpretation of the usage patterns associated with that cluster.
Table 25 The usage patterns for each of the k-means identified clusters.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>K = 3</th>
<th>K = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_M_U</td>
<td>Minimal Usage Pattern (K = 3)</td>
<td>Minimal Usage Pattern (K = 4)</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>C_J_U</td>
<td>Just-in-Time Usage Pattern (K = 3)</td>
<td>Just-in-Time Usage Pattern (K = 4)</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>C_E_U</td>
<td>Excited Usage Pattern (K = 3)</td>
<td>Excited Usage Pattern (K = 4)</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>C_C_U</td>
<td>Consistent Usage Pattern (K = 4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Graph" /></td>
<td></td>
</tr>
</tbody>
</table>
Table 26 The number of students assigned to each usage cluster when K = 3 and K = 4.

<table>
<thead>
<tr>
<th>Cluster Assignments</th>
<th>K = 3</th>
<th>K = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_E_U</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>C_J_U</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C_M_U</td>
<td>40</td>
<td>29</td>
</tr>
<tr>
<td>C_C_U</td>
<td></td>
<td>12</td>
</tr>
</tbody>
</table>

Table 27 The centroid values for each of the features (unitized event counts per week).

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_M_U</td>
<td>3</td>
<td>0.03</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.03</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>C_J_U</td>
<td>3</td>
<td>0.31</td>
<td>0.13</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.27</td>
<td>0.14</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>C_E_U</td>
<td>3</td>
<td>0.01</td>
<td>0.71</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.03</td>
<td>0.89</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>C_C_U</td>
<td>4</td>
<td>0.03</td>
<td>0.32</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

7.4.3.2 Observed Usage

The activities that we encouraged students to perform using VocabNomad were challenging, which is evidenced by observations of students working on them in groups outside of class. This is also supported by Sensei_4’s report of overhearing some of the students complaining, in Japanese, about how hard the VocabNomad-based worksheets were. That said, an excerpt from a worksheet that Nihon_S37 left behind demonstrates that the task was within their abilities even if they did not always derive the correct answer (Figure 30). Both of the worksheets that were left behind by Nihon_S37 can be seen in their entirety in Appendix J.

Students seemed to enjoy the text-to-speech feature and would occasionally play around with it during class. In one case, a student accidentally triggered a rather long vocabulary entry while a

Figure 30 Nihon_S37’s response for one of the items from a VocabNomad-related worksheet. He translated handout as print out (purinto).
teacher was speaking and then kept touching the screen to try to get it to stop speaking. Evidence of their use of the text-to-speech feature is also present in the log files (i.e., Listening to a Sentence row in Table 24). Since we rarely heard the use of text to speech during class, it is likely that the majority of the logged usage of this feature occurred outside of class.

Nihon_S44, who appears to have performed minimal personalization on his device, asked if it was okay to edit the sentences associated with vocabulary items. I told him that it was and he proceeded to edit the sentence. This indicates that some students perceived the tool as a knowledge transmission device rather than a knowledge construction tool, especially if we consider the limited number of content creation and editing events (Table 24).

Some students consistently brought their devices to class where they were seen sitting on student desks. Others only brought their devices some of the time. Few students were observed using their electronic translators during EI2. However, the electronic translators were seen sitting on their desks during their other English classes, where they were observed using them.

7.4.4 Morphological Knowledge

The measures of morphological knowledge are used to describe student abilities rather than determine student learning outcomes. The reliability measures for the full morphological structure and morpho-orthographic choice tasks were low so item analysis was conducted and unreliable items were removed. This resulted in 17 of the 35 items being retained for the morpho-orthographic choice task, an \( \alpha \) of .63, and a Guttman split-half reliability measure of .64. Item analysis reduced the structure test to 21 items from the original 28. This reduction in test items resulted in an \( \alpha \) of .77 and a split-half reliability of .78. As expected, the unbiased alternate forms reliability of the measures of morphological knowledge (\( r(52) = .47, p < .001 \)) showed a moderate relationship between these measures. These reliability measures give us a sense of how consistently each of the tasks measured student morphological knowledge: ideal values would be 1.0. However, the observed reliability of these measures was sufficient for the purposes of this study.

Student scores (Table 28) on Carlisle’s morphological structure test were normally distributed for both the treatment (\( p = .79 \)) and control groups (\( p = .16 \), with Nihon_S13 being identified as
Table 28 Student scores (%) on morphological measures.

<table>
<thead>
<tr>
<th>Task</th>
<th>Treatment</th>
<th></th>
<th></th>
<th>Control</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>Min.</td>
<td>Max.</td>
<td>M (SD)</td>
<td>Min.</td>
<td>Max.</td>
</tr>
<tr>
<td>Morpho-orthographic Choice</td>
<td>78.1 (13.97)</td>
<td>41.2</td>
<td>100.0</td>
<td>79.0 (13.52)</td>
<td>58.8</td>
<td>94.1</td>
</tr>
<tr>
<td>Morphological Structure Test</td>
<td>46.3 (17.16)</td>
<td>0</td>
<td>81.0</td>
<td>30.3 (7.86)</td>
<td>21.4</td>
<td>40.5</td>
</tr>
</tbody>
</table>

an outlier. Treatment-group scores for the morpho-orthographic choice task (Table 28) were not normally distributed ($p = .04$) even though control group scores were ($p = .40$), as measured by a Shapiro-Wilk test. One outlier (Nihon_S04) was identified from the treatment group for the morpho-orthographic choice task.

Outliers were excluded when performing group comparisons. There was no significant difference in student performance on the morpho-orthographic choice task ($U = 151.50$, $p = .74$) indicating that both groups were similarly good at recognizing correctly spelled words. However, there was a significant difference in student performance on the morphological structure test ($t(51) = 2.76$, $p < .01$) based on group membership with equal variances assumed ($F = 2.29$, $p = .14$). This indicates that students in the course where VocabNomad was deployed were more able to modify words so that the words could be used in different settings.

An analysis of student errors was performed and was used to inform the reports that were given to students and the school. It was also used to select the activities that were provided in student workbooks. For more information on the content of these documents please see Appendix J and 7.9 Stakeholder Reports.

Both groups exhibited negative transfer from their L1 (Japanese) to their L2 when they selected *complition* over *completion* during the morpho-orthographic choice task: the romanji *i* is pronounced similarly to the *e* in *completion*. Both groups also had trouble correctly identifying when a consonant would be softened as the result of the addition of a suffix (e.g., when *tion* is appended to *absorb* in *absorption*). Students struggled with determining when consonants should be doubled, which may be a result of negative transfer since consonant doubling in Japanese does not affect the pronunciation of vowels. Negative transfer may also contribute to student errors when they were uncertain about dropping vowels (e.g., *true* to *truly*).
Table 29 PPVT-4 reliability statistics by administration.

<table>
<thead>
<tr>
<th>Reliability Measure</th>
<th>Start</th>
<th></th>
<th>Middle</th>
<th></th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach’s α</td>
<td>.995</td>
<td>.83</td>
<td>.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guttman Split-half Coefficient</td>
<td>.88</td>
<td>.81</td>
<td>.70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Both groups struggled with modifying words that required the root to be changed (e.g., long to length). Morphologies that required the addition of th were also difficult for students (e.g., four to fourth) as were items that required a noun to be placed in the blank space.

7.4.5 PPVT-4

Like with the assessments of morphological knowledge, item analysis was performed on student responses to the PPVT-4 and unreliable items were removed. The reliability measures for each administration (Table 29) meet minimum reliability requirements but are lower than those reported for the standard administration method (Dunn & Dunn, 2007).

Participant scores on the PPVT-4 were, at a minimum, weakly correlated with the average of their performance on the assessments of morphological knowledge for the first \((r(45) = .44, p < .01)\), second \((r(52) = .54, p < .001)\), and third \((r(52) = .33, p = .02)\) administrations of the test. The first \((r(45) = .32, p = .02)\) and second \((r(52) = .38, p = .02)\) administrations of the PPVT-4 were also weakly correlated with students’ self-assessments of their English-language abilities.

Participant scores on the PPVT-4 (Table 30 and Figure 31) were normally distributed for both the treatment \((p = .31)\) and control \((p = .33)\) groups, as measured by Shapiro-Wilk.

Since the assumption of sphericity was not violated by the treatment group’s scores \((\chi^2(2) = 0.22, p = .89)\), a repeated measures ANOVA was used to test for differences in the treatment group’s scores for receptive vocabulary knowledge. A statistically significant difference in treatment group PPVT-4 scores (Table 30) based on time of administration \((F(2) = 23.44, p < .001)\), partial

Table 30 PPVT-4 percentage scores by administration.

<table>
<thead>
<tr>
<th>Group</th>
<th>Start</th>
<th></th>
<th></th>
<th>Middle</th>
<th></th>
<th></th>
<th>End</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>Min</td>
<td>Max</td>
<td>M (SD)</td>
<td>Min</td>
<td>Max</td>
<td>M (SD)</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>Treatment</td>
<td>56.84 (6.5)</td>
<td>31.59</td>
<td>72.37</td>
<td>68.72 (12.8)</td>
<td>12.50</td>
<td>92.50</td>
<td>62.96 (12.1)</td>
<td>36.67</td>
<td>83.33</td>
</tr>
<tr>
<td>Control</td>
<td>46.00 (5.2)</td>
<td>55.00</td>
<td>87.50</td>
<td>44.14 (2.7)</td>
<td>40.00</td>
<td>76.67</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
$\eta^2 = .35$ was found. Post-hoc paired t-tests with Bonferroni correction indicated a significant improvement in the vocabulary knowledge of the treatment group during phase 1 ($p < .001$) but not during phase 2 ($p = 1.0$).

Phase 1 is when participants were using VocabNomad moderately heavily which indicates that their improvement is linked to their VocabNomad usage. The fact that the final PPVT-4 score for this group was significantly higher than that obtained during the first administration ($p < .01$) indicates that the improvement in student vocabulary knowledge was not temporary.

No significant change in the control group’s scores was observed between test administrations ($t(6) = 0.96, p = .38$) as measured by a paired-samples t-test.

T-tests assuming equal variance showed no significant difference in student PPVT-4 scores at the mid ($t(52) = 0.20, p = 0.84$) or final administrations ($t(52) = 0.22, p = .83$) based on group membership.

The results of the within and between group tests of student performance on the PPVT-4 support the argument that students learned when using VocabNomad.
Table 31 Distribution of vocabulary scores from the final exam.

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Q1</th>
<th>Mdn</th>
<th>Q2</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>30</td>
<td>70</td>
<td>88</td>
<td>96</td>
<td>100</td>
</tr>
</tbody>
</table>

7.4.6 Final Exam Vocabulary Scores

Treatment group scores on the vocabulary portion of their final exam were not normally distributed ($p < .001$), likely because of a ceiling effect. Table 31 shows the score distribution, which is heavily skewed with three quarters of the class answering at least 70 percent of the questions correctly. Students’ average PPVT-4 scores were not correlated with their final exam vocabulary scores ($p = .18$) nor were their self-perceived English language abilities ($p = .18$). However, their morphological knowledge was weakly correlated with their final exam scores ($r(47) = .40, p < .01$) as was their overall English language ability when all other language assessments were combined ($r(47) = .33, p = .02$).

7.4.7 Self-Assessed Language Ability

The questionnaire used to collect student self assessments of their language abilities was sufficiently reliable for measuring their proficiency in English ($\alpha = .91$) and Japanese ($\alpha = .89$).

Since the sample size was relatively small, the Shapiro-Wilk test for normality was used. The results of the test for each administration of the Japanese measure can be seen in Table 32. Table 32 indicates that the self-report data was not normally distributed for the control group at both administrations, and a Related-Samples Wilcoxon Signed Rank Test showed no significant difference ($Z = -0.68, p = 0.50$) in the control group’s self-reported Japanese-language abilities (see Table 33). The treatment group’s self-reported Japanese-language abilities also violated the

Table 32 Shapiro-Wilk test p-values for participant self-assessments of their Japanese-language (J) abilities.

<table>
<thead>
<tr>
<th></th>
<th>Control - Japanese</th>
<th>Treatment - Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Speak</td>
</tr>
<tr>
<td>Start</td>
<td>0</td>
<td>.00</td>
</tr>
<tr>
<td>Mid</td>
<td>7</td>
<td>.10</td>
</tr>
<tr>
<td>End</td>
<td>7</td>
<td>.06</td>
</tr>
</tbody>
</table>

Speak – speaking, Und. – understanding, Read – reading
### Table 33 Student self-assessed Japanese-language (J) proficiency.

<table>
<thead>
<tr>
<th></th>
<th>Control - Japanese</th>
<th>Treatment - Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Speak (SD)</td>
</tr>
<tr>
<td>Start</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td>7</td>
<td>8.69 (1.2)</td>
</tr>
<tr>
<td>End</td>
<td>7</td>
<td>8.86 (0.9)</td>
</tr>
</tbody>
</table>

Speak – speaking, Und. – understanding, Read – reading

assumption of normality (Table 32). No significant differences in treatment group Japanese-language ability were measured using a Friedman ANOVA by ranks test ($\chi^2(2) = 4.41, p = .11$).

This suggests that each group’s self-perception of their Japanese-language ability remained unchanged, which is what should be observed for student self-assessments of their mother tongue.

As expected, an independent samples Mann-Whitney U-Test indicates that there was no difference in student self-perceptions (see Table 33) of their Japanese-language abilities based on group membership at the mid ($U = 160.50, p = 0.92$) and end ($U = 136.50, p = 0.47$).

The Shapiro-Wilk test indicated that the English-language self-assessments that were performed by the treatment group were normally distributed (Table 34), and the assumption of sphericity was not violated ($\chi^2(2) = 4.46, p = .11$). A repeated measures ANOVA showed a difference approaching significance based on time of administration ($F(2) = 2.57, p = .08, \eta^2 = .50$).

To determine whether treatment group self-perceptions had changed based on study phase, a paired samples t-test of the score gains was conducted (see Figure 32): treatment group scores improved significantly more during the first half of the term (phase 1) ($t(46) = 2.01, p = .05$).

### Table 34 Shapiro-Wilk test results for participant self-assessments of their English-language (E) abilities.

<table>
<thead>
<tr>
<th></th>
<th>Control - English</th>
<th>Treatment – English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Speak</td>
</tr>
<tr>
<td>Start</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td>7</td>
<td>.16</td>
</tr>
<tr>
<td>End</td>
<td>7</td>
<td>.33</td>
</tr>
</tbody>
</table>

Speak – speaking, Und. – understanding, Read – reading
The control group’s English-language self assessments were also normally distributed (Table 35) and no significant difference was found between the control group’s mid and end point self-assessments ($p = 1.0$) as measured by a paired samples T-Test.

Independent samples T-Tests, with equal variances assumed, showed no significant difference between the self-assessed language abilities of each group at the end of the study ($t(52) = 1.57, p = .12$). However, the same test revealed that the control group’s English-language abilities were significantly lower than the treatment group’s at the mid-point administration ($t(52) = 1.97, p = .05$). The assumption of equal variances is supported by Levene’s test for the end ($p = .90$) and

<table>
<thead>
<tr>
<th>Table 35 Student self-assessed English-language (E) proficiency.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control - English</strong></td>
</tr>
<tr>
<td><strong>n</strong></td>
</tr>
<tr>
<td>Start</td>
</tr>
<tr>
<td>Mid</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>End</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Speak – speaking, Und. – understanding, Read – reading
mid \( (p = .88) \) point self-report data.

7.4.8 The Relationship Between Presentation Groups and Measures

There were no significant correlations between students’ membership to their presentation group and any of the measures. This indicates that the language abilities of each group were relatively balanced. However, there was a weak relationship \( (r_s(45) = .25) \) approaching significance \( (p = .09) \) between student group membership and the system usage patterns that had been identified using k-means clustering with K set to 4. This would be expected since student activities are likely to influence one another when working in groups towards a common goal. This statistic also supports modeling theories that describe how one’s actions are influenced by the actions of his or her peers (Bandura, 1986; Pajares, 2002).

7.4.9 English Language Abilities

In addition to building models of system usage, models of student language proficiency were built. These models helped to group students into classes of different abilities and served to reduce the dimension space for later model building.

Once again, RapidMiner 5.3.015 was used. K-means clustering was applied to the unitized averages of student morphological knowledge, PPVT-4 scores, and self-assessed English ability. K values of 2, 3, 4, and 5 were attempted. The maximum number of optimization steps and runs were set to 1000 and a random seed was used.

Visual inspection of the centroid plots and cluster assignments led to the rejection of the clusters that were created using 2 or 5 for K since they did not add meaningful information to the clusters that were formed when K was set to 3 or 4. The DBI is 0.13 when K = 4, indicating that this was a reasonable choice since its DBI is lower. With K set to 4, the following language ability clusters were identified:

- Moderately strong performers with realistic self-perceptions (C_MSR_E)
- Weak performers with realistic self-perceptions (C_WR_E)
- Moderately strong performers with low self-perceptions (C_MSL_E)
- Moderate performers with realistic self-perceptions (C_MR_E)
Table 36 The number of students assigned to each language-ability cluster when K = 3 and K = 4.

<table>
<thead>
<tr>
<th>Cluster Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_MSR_E</td>
</tr>
<tr>
<td>K = 3</td>
</tr>
<tr>
<td>K = 4</td>
</tr>
</tbody>
</table>

If K is reduced to 3 then the DBI increases to 0.15, and we no longer see the moderately strong performers with low self-perceptions (C_MSL_E). The number of students that were assigned to each cluster can be seen in Table 36.

As might be expected, the cluster assignments for English language ability were weakly correlated with students’ final exam performance ($r_s(45) = .30, p = .04$) when K = 4. There was an additional weak relationship ($r_s(45) = .31, p = .04$) between student exposure to English through media or English-language environments and their language-ability cluster membership when K = 4.

We can go beyond the machine-generated classifications of students based on quantitative measures of their language abilities and consider some of the work that they performed during class. This work was not submitted but revealed some of the challenges that students were facing and can be used to characterize their language skills. When preparing their presentation slides, students were often observed making subject-verb agreement errors. They also seemed to struggle with making appropriate word choices, especially with respect to the collocations that they chose. Some students were able to identify or correct their own errors if their attention was drawn to the error, especially when they had accidentally used a homophone (e.g., two and to). Students also seemed to have trouble with mapping Roman alphabet characters to their English sounds as is evidenced by their inconsistency in spelling their own names. This is likely a result of negative transfer from using the Romanji alphabet for Japanese words.

7.4.10 The Relationship Between Usage and Language

The usage-based clusters (K = 3) and participant English ability clusters (K = 4) were weakly correlated ($r_s(45) = 0.37, p = .04$). The usage-based clusters (K = 3) were also weakly correlated with students’ average PPVT-4 scores ($r_s(45) = .36, p = .01$), their level of Exposure to English
(r_s(45) = .39, p < .01), and their overall English ability (r_s(45) = .32, p = .03), which includes all language measures except the final exam.

Given these relationships and the existence of identifiable clusters of system usage and students’ English-language ability, a discovery with models approach was taken (Demmans Epp, 2015). This method uses the output from one model as input to another model and is a relatively recent addition to those used to analyze data from educational systems (Greer & Mark, 2015).

W-JRIP was applied to the usage-based and language-ability clusters to try and predict each student’s score on the vocabulary section of the final exam. The default parameter settings for W-JRIP with leave-one-out cross-validation were used. The model that uses K = 4 for usage-based clusters and K = 3 for language-based clusters resulted in a classification accuracy of 72.34 percent (chance is 44 percent) and a kappa of .258. This does not show a strong relationship but provides a starting point from which we can evaluate later learner usage of the system (Figure 33). It also provides us with a rule that could be used for prediction and system adaptation. This rule (Rule 1) predicts the group membership for students whose scores fall in the range between 60 and 80 percent. Rule 2 states that all other students should be assigned to the high-performer group, which overfits to the exam results that were highly skewed as a result of a ceiling effect. However, this rule could be used and later refined following the collection of additional data.

Should we want to predict student score gains on a more reliable measure of their vocabulary knowledge we can build a model that predicts how much their PPVT-4 score is expected to change given their self-assessed English-language proficiency, their morphological knowledge, and their system usage patterns.

**Rule 1:** (C_WR_E) and (C_J_U or C_E_U) => FinalExamPerf = Middle Performers
**Rule 2:** => FinalExamPerf = Top Performers

*Figure 33 Decision rules predicting student final exam performance.*

---

10 The modelling approach used here was reported in (Demmans Epp, 2015). See Appendix L for a complete list of publications and their relationship to my thesis work.
W-J48 was used for this purpose with students’ average self-assessed English proficiency, the usage-based clusters that had been determined using $K = 3$, and the average of students’ morphological knowledge as inputs to the classifier. Students’ PPVT-4 score gain from the first half of the term was used as the dependent variable.

This results in a decision tree with a classification accuracy of 61.70 percent and a kappa of .312. The resulting decision tree (Figure 34) predicts that students who have the best self-assessed language abilities will also experience the highest score gain. For those with lower self-assessed English proficiency, their morphological knowledge plays a larger role in predicting their PPVT-4 score gains. It is worth noting that student usage patterns do not play a role in the decision tree unless the usage clusters that were assigned when $K = 4$ are used, in which case the classification accuracy is 57.45 percent and kappa is .250. Both trees start with the same decision point that is based on student self-perceptions. However, the $K = 4$ tree is able to employ student usage patterns to help predict how students will perform.

Looking at the confusion matrices (Table 37), we can see that both models perform better than chance (0.44 when $K = 3$; 0.43 when $K = 4$) and have similar structures (Figure 34), which may mean that the use of the $K = 4$ tree would be more prudent within an adaptive system since it probably underperforms as a result of class imbalances in the sample data and could improve over time, especially if we consider the relatively small data set on which the model was built. It is also more consistent with the results of the W-JRIP model that was used to predict student performance on their final exam. These results indicate that the approach used might provide better results if additional studies with more students were conducted.

| Table 37 The confusion matrices for the W-J48 Decision trees. |
|---------------------------|---------------------------|
| **K = 3** | **K = 4** |
| | | True | Precision (%) | | True | Precision (%) |
| | | High Med Low | | | High Med Low | |
| Predicted | | 0 1 1 | 0.0 | | 0 2 2 | 0.0 |
| | Med | 3 8 2 | 61.5 | | Med | 3 7 2 | 58.3 |
| | Low | 5 6 21 | 65.6 | | Low | 6 2 20 | 64.5 |
| Recall (%) | 0.0 53.3 87.5 | | Recall (%) | 0.0 46.7 83.3 |
Figure 34 W-J48 Decision trees that predict a student’s PPVT-4 score gain category. The tree for K = 3 is on the left, and K = 4 is on the right.

7.5 Barriers and Challenges Faced

WiFi connection problems prevented some students from using the on-demand vocabulary creation or sharing vocabulary during the first half of the term. However, these features were made available during phase 2 after purchasing a router that allowed upwards of 50 connections at a time.
Another barrier was students’ inability to turn the text to speech feature off mid sentence even though the interface allowed it. The embarrassment caused by activating the text to speech when a teacher was speaking may have discouraged some students from using the application. This situation could be overcome with training, further system exploration, or, as suggested by Sensei_2, the use of headphones where only one ear bud is used. The suggested use of headphones matches with what some teachers have done in special education settings when their students were using mobile communication support tools (Campigotto et al., 2013).

Limited classroom time meant that measures of student cognitive abilities and phonological skills could not be taken. Time restrictions that related to students not following instructions precisely or quickly also led to the adaptive testing being cancelled, which meant that more detailed tracking of student vocabulary learning was not possible.

The lack of teacher trust in students’ ability to create learning materials limited student languaging activities and system use. When this is combined with the teacher’s development of vocabulary and the pre-loading and distribution of support materials through VocabNomad, it may partially explain why students did not edit vocabulary entries even though this feature was demonstrated to them. These teacher actions may have communicated that students were not welcome to add new vocabulary and contributed to the lack of content creation by students. As indicated by Sensei_1, it could be the result of insufficient training or scaffolding in system use and, more specifically, in the creation or editing of content.

Students’ highly variable oral and aural abilities present challenges when designing structured learning activities with the aim of ensuring that languaging or an increased depth of processing occur. Sensei_2 suggested providing additional L1 support based on each learner’s skills. This support could then be removed as their language proficiency increases, which could reduce the barrier to entry that some students may have experienced as a result of their limited English proficiency.

Sensei_1 felt that student interest in using the application waned over the course of the term, which is reflected in the system usage logs. However, the existence of the just-in-time (C_J_U) and consistent (C_C_U) usage clusters indicates that some students did not lose their motivation or were willing to use the system even after their motivation had waned provided that the system
addressed a need. This indicates that a lack of interest can be partially overcome through the use
of structured learning activities that support system use.

7.6 Successes

Sensei_1 felt that having students learn English through English was one of the major successes
of this project. Using English to learn English was considered a success because students had
previously only learned their L2 through the translation of their L1.

The student creation of vocabulary lists for their classmates was considered a success. This
creation was performed on paper, which partially explains lower usage levels and a lack of
content creation or editing events in the second half of the term (Table 23 and Table 24).
Teachers at the school think that it may have been the first time that this type of activity was
done in their English classes. The addition of teacher curation to this process helped ensure that
errors would not be propagated when the word lists were shared and used by students. It is
believed that the curation process contributed to the success of the task since it meant that
students took it seriously.

In addition to student creation of vocabulary lists and their attempts to use English materials to
learn English, student performance on the vocabulary portion of their final exam indicates that
they learned many of the materials that were distributed through VocabNomad.

7.7 Opportunities

Interactions that focused on the technology seemed to help students overcome social barriers and
enabled them to interact with me using English. Nihon_S45 used me as one of his conversation
partners when I was observing another class that he was in, which was somewhat surprising to
his regular classroom teachers (Sensei_1 and Sensei_5) since I was, largely, considered an
outsider. Other students, from the course (EI2), would ask me for help when they were working
on essays or their presentation slides during their spare period, one student approached me to
discuss his music preferences, another (Nihon_S33) discussed Taylor Swift with me, and one
(Nihon_S11) decided that it was his job to teach me about Japanese culture or otherwise
entertain me. This provided students with additional opportunities for languaging and noticing.
With the exception of Nihon_S33, this additional languaging may have helped their English
learning. It is unlikely that Nihon_S33’s English was affected through his interactions with me because of his already high English proficiency.

Students approaching me and teacher reactions to these events support Sensei_1’s observation that students do not normally learn English through English and that this application enabled that process in a safe way. This approach to English learning could be further supported through additional activities such as those reported by Wong and Looi (2010), where students photographed and shared situations that they encountered in their real lives.

Student creation of lists of vocabulary with accompanying sentences was a big step towards more socio-collaborative approaches to language-learning. The next step would be to have students supply all of the content for a vocabulary entry. After this, one might be able to have a class curate its own content using an explicit curation model rather than the implicit model that currently exists within the community of users through their sharing and editing of content.

Providing students with additional feedback, adaptive support, and assessment opportunities through the application may help increase student application usage. Providing specific targeted tasks during the more open-ended, second half of the term may have encouraged students to focus more on their language learning activities within the system. Sensei_1 suggested that we go one step further by removing the paper-based supports, if the training were to emphasize the L1 support that is available through VocabNomad or the device. The removal of the paper-based resources would allow teachers to give students more feedback based on their activities since more of the students’ activities would be captured through the system. Sensei_1 had wanted to provide students with feedback about their system-usage activities throughout the study but this was not done because I did not want to impose a specific learning path on students. Now that we know that VocabNomad can support both student learning and their learning activities, it could be appropriate to remove this additional support and give students feedback based on their system usage.

In addition to these logistical opportunities, Sensei_2 indicated that the tool would be useful in themed reading courses or a Test of English as a Foreign Language (TOEFL) preparation course since these courses place considerable emphasis on vocabulary development.
7.8 Discussion

We first discuss patterns in the results from the quantitative measures. This is followed by a
discussion of the research questions that were asked in this chapter as well as how this study’s
results tie into the larger research questions that are addressed throughout this thesis. Following
from this, student responses to the study are presented as are the implications that this study
holds for educators who find themselves in similar contexts.

7.8.1 Measures

The treatment group’s improved PPVT-4 score is noteworthy given the measure’s stability and
the conditions under which the test was written. Students had just completed a midterm exam for
another subject. While this exam may have heightened their focus, their behaviour on the day
and apparent lack of energy indicates that they were tired and may not have been performing
their best when writing the exam. The non-significant decrease in PPVT-4 scores between the
mid and end-point administrations may be the result of measurement error or student regression
towards the mean.

The EI2 final exam was similar to the midterm in many respects, except that it contained a large
vocabulary section. Students performed well on this portion of the exam and ceiling effects were
observed. These ceiling effects prevent the development of better models of student vocabulary
learning based on their system usage because test scores do not represent the full range of student
abilities. The system’s adaptive testing might have helped prevent this problem had it not been
cut from the study for logistical reasons.

The control group’s self-assessment results may be due to selection mortality as a result of drop
out (6 of the 13 chose not to complete the study). Whereas the non-significant decrease in the
treatment groups’ scores during phase 2 may be the result of regression towards the mean or
measurement error. It is also possible that the midterm scores of treatment-group students
negatively affected their self-perceptions since the listening task was particularly difficult.
Alternatively, the delivery of presentations without being allowed to use a script may have
decreased student confidence in their abilities or increased their awareness of their skills.
Selection bias may account for the lack of difference between the performance of the control and treatment groups on most measures. The students who opted in to the control group were all highly motivated even if the school had not considered them high performers. Their motivation and willingness to perform extra learning activities may explain why they exhibited similar performance on most measures at the mid-point administration. The school’s decision to group these students based on their abilities is partially supported by the results of the production-based morphological structure test and students’ self-assessed language proficiency at the mid-point. However, the difference in self-assessed ability could be the result of the testing and student streaming programs that are used within the school.

7.8.2 Research Questions

The results support H1. Student receptive vocabulary (i.e., PPVT-4 scores) increased during the first phase of the study, which was when students were using VocabNomad more heavily (RQ1). A ceiling effect was also observed on the vocabulary portion of their final exam which involved performing receptive and productive vocabulary activities using the words that had been distributed via the application. These test results indicate that students learned the targeted words and that their vocabulary knowledge increased while using the tool (RQ1).

The amount of support that VocabNomad provides may be reflected in its usage since students only exhibited a score gain for their self-assessed English-language abilities or general vocabulary knowledge (PPVT-4) when they were using the application the most. This is when students were performing structured activities using the application. The control group did not exhibit gains in their vocabulary knowledge or self-assessed language abilities. So, neither group improved their vocabulary knowledge or their English language abilities when they were not using the application.

Student vocabulary knowledge correlated with their system usage patterns (RQ5). This correlation revealed a weak relationship between PPVT-4 scores and usage patterns. The additional weak correlation between student system usage and their language abilities helps to define the relationship that exists between ELL vocabulary knowledge and the usage of tools like VocabNomad (RQ3). The predictive rules and decision trees that were developed define this relationship and are consistent with previous research (Nagy, 2007; Wagner et al., 2007): they
include morphological knowledge as a predictor of vocabulary knowledge. In addition to the consistency of the predictive models with prior research, the descriptive models of system usage patterns are consistent with recent work that showed the existence of the minimal activity cohort (C_M_U) in a Japanese context and a version of the excited users (C_E_U) or just-in-time users (C_J_U) in a Taiwanese context (Stockwell & Liu, 2015).

The predictive rules that were generated using W-JRIP accurately identify the mid-performing students. However, they misclassify low performers as high performers. While this maintains the school’s status quo by keeping students in the top performing group even if they are no longer top performers, it poses a risk to their learning. Therefore, a system that incorporates these rules should most likely use a human-in-the-loop approach. Teachers (the humans in the loop) could then help the system to distinguish between true high performers and low performers so that the low performers get the support that they need. This can save teacher effort by focusing it where it is most needed.

Student responses to the English self-assessment questionnaire indicate that their abilities to function in English by speaking, reading, or listening improved when they were using the application most heavily. This supports VocabNomad’s ability to advance student communication by improving their underlying speaking and listening skills (RQ1).

The improvements in student abilities may be due to increased time on task or the increased depth of processing that some students may have experienced when completing the optional structured activities that were encouraged during phase 1. While it might have been expected that the vocabulary creation task that students were assigned in phase 2 would have resulted in student languaging activities or a greater depth of processing, this does not appear to have been the case since learning was not observed. We do not know why this happened but two possibilities exist. Students may have used a service such as Google translate or Google search to create or find the sentences that they included on their vocabulary lists. The other option, which seems more likely given the errors that were present in some of the sentences, is that students were strategic and chose to use sentences that required knowledge with which they were already comfortable.
We successfully integrated the application into classroom activities as a support tool (RQ4). This was more easily done during the first half of the term (phase 1), when we were trying to build student knowledge and skills, than it was during the second half of the term (phase 2), which was when students were performing self-guided study in groups. This self-guided study was used to support their research into a selected topic so that students could present on that topic. Students used the application to support supplementary and group activities that were performed in class and out of class. However, their use of the application during the second half of the term was far more sporadic and far less structured.

The use of other tools, such as webiblio, highlights how other MALL tools can be used to complement the approach that is provided through VocabNomad (RQ1). Allowing access to these tools alongside the use of structured activities and worksheets helped ensure VocabNomad’s integration into the course. The application’s integration during phase 1 was further supported by teachers encouraging students to scaffold their small group work by using VocabNomad or the worksheets that they had completed (RQ2). Students’ low reliance on the handheld translators that they normally use in class indicates that students may not need the level of L1 support that many teachers from this school believe is necessary as long as appropriate scaffolding is provided through other avenues, such as VocabNomad, other MALL tools, or worksheets that provide cognitive support for previously covered materials.

7.8.3 Student Responses

The observed lack of over-confidence in students’ self-perceptions of their abilities is likely a cultural phenomenon. Japanese culture emphasizes humility, which may have contributed to students having realistic or low self-perceptions with respect to their English language abilities. The school’s use of extensive and repeated testing may have influenced students’ self-perceptions. These two factors contributed to the absence of certain English-language clusters that may have been expected in other cultural settings. Among them would be the existence of over-confident, low-performing students. So, while the clusters that were identified in this study are expected to generalize across cultures, they are in no way a complete set.

Students in this study, like those in the MyVoice study (Chapter 3), did not seem to perceive VocabNomad as an application over which they could exert complete agency. This is evidenced
by a student asking if it was okay to change a sentence as well as the lack of editing performed by students. Part of this may be due to transfer from other applications that are largely content-delivery mechanisms even if they are interactive, as is the case with games. Another factor that may have contributed was insufficient training and encouragement for adding or editing content. When this subtler cue is combined with the teacher creation of content and the teacher curation of student-generated content, it is possible that we essentially told students that they were not supposed to modify or add content.

7.8.4 Implications for Educators

Two of the major successes from this study were that students in this context could learn English through English and they could create learning materials for their classmates. Prior to this, these students had primarily learned English through translation. This project required students to approach their language learning differently and student creation of learning materials was considered innovative by teachers in this setting.

Some students added words that had already been covered in this course or were being covered in their other English courses. This repetition, whether intentional or accidental, may have helped students to learn the words. It may reflect students’ belief about their classmates’ knowledge of the material or a strategic move that acknowledged the word’s importance due to its reappearance between courses or topics. The repetition of some of these words may also be the result of students attempting to meet the teacher set quota for the minimum number of words that had to be submitted. There were presentation groups that could have submitted longer word lists to better support their classmates’ understanding of the presentation topic but who chose to submit the number of words that had been suggested by the teachers. I would recommend setting an expectation range rather than providing a specific number. This will allow students to submit the words that they think their classmates do not know while still giving students some freedom. This level of freedom and willingness to create and share learning content is supported by the actions of the students who added, edited, or shared vocabulary through the MALL tool.

Our data confirms claims that students need training in MALL tool use and that activities need to progress in difficulty based on student technological knowledge and language knowledge (Stockwell & Hubbard, 2013; Sweeney, 2013). While we trained students in the use of the
application, it is likely that they needed additional training at intervals throughout the term. Delivering repeated training that progressed with student knowledge of the application would have provided an opportunity for students to remain involved in the process. It would have also allowed students to help their fellow learners by providing each other with content. More importantly it could have combated existing cultures of MALL use and the impression that students were not allowed to edit or create materials unless they were explicitly told to perform these activities.

Student creation of content for one another may need additional monitoring or curation procedures to be enabled. The practice of having students create content for each other fits with a more constructivist view of learning and may need additional scaffolding in environments that have not adopted this approach to learning. This need for scaffolding may be independent of the technical support for enabling collaborative content creation and sharing. Our use of small incremental increases in learner autonomy enabled this process in the classroom where we deployed the adaptive MALL tool. Larger steps may be possible in classrooms where more socio-collaborative approaches are already used.

If we want to see student learning, it is not enough to introduce a tool into the classroom. It must be effectively integrated. In many classroom settings, structured activities will need to be planned. These activities should be challenging and designed in a way that increases student depth of processing. Students need to be able to work with the meaning of words effortfully if they are going to develop more extensive knowledge of their meaning and use. The use of structured activities can be slowly removed. However, the use of the tool as a support for general classroom activities will need to be modeled so that all students can use it effectively. In settings where a more constructivist approach is used, having students show each other how they have used the application could be helpful. In our case, students did this on their own outside of class. However, younger students may need more support to accomplish these types of activities.

7.9 Stakeholder Reports

One of the steps in action research is to reflect over your practices and adjust them with the intent of improving a weakness that has been identified (Mills, 2007). As part of this process and my departure from the research site, I provided the participating students and the school with
reports highlighting different aspects of the project. Student reports highlighted their test results and provided students with English-language activities to help develop their skills. The school report focused on the strengths and weaknesses of the group as a whole as well as some of the lessons that we learned should the school want to implement a similar program in the future.

7.9.1 Student Reports

Student reports and activity books were distributed through their homerooms during the last week of the term. These reports contained some basic information about the tests that they had written. The reports provided basic descriptive statistics of the group’s performance in the form of a box plot. Each student’s score was included and marked on the box plot. Only the score of the student who was receiving the report was provided. That is, Nihon_C01 would only see his score alongside the distribution of scores as it was represented through the box plot. The template that was used for the reports can be seen in Appendix J.

Additional feedback was provided in a comments section at the end of the report. This feedback was qualitative in nature and included tips that were adjusted based on each student’s test performance, co-curricular activities, and classroom behaviours. For some students this meant encouraging them to talk with others in English, while other students may have been encouraged to read more or to start signing out library books with higher grade level ratings.

These reports were accompanied by booklets containing activities that targeted skills that students showed difficulty completing. This included several worksheets that were dedicated to the meaning and use of prefixes and suffixes. The booklets included worksheets that asked students to derive the meaning of words by decomposing them into their constituent parts or through other contextual cues, such as the surrounding text. The answer key and supplemental materials were left in the English department so that students could check their work and access more resources should they want to continue improving their English.

7.9.2 School Report

The report can be seen in Appendix J. It is a high-level overview of the results of the study. The school’s report showed the box plots for the performance of all students on the PPVT-4, morpho-orthographic choice task, Carlisle’s morphological structure test, and the student’s self-assessed
English language abilities. The report also briefly discusses student usage of the application and its relationship to student performance on these measures. Information about some of the types of errors that students made when completing these measures was provided and the report suggests certain types of learning activities that could be of benefit to students.

7.10 Limitations

Since the PPVT-4 is a standardized test, it may not have measured the full change in student vocabulary knowledge because it did not test the vocabulary that they were studying but rather the larger construct of American-English vocabulary knowledge. Its reliability was slightly reduced as a result of the administration method. The order of the subtests may have affected the results since each subtest may have had different difficulty levels. However, it allowed one test to be used many times and reduced the potential for test-item familiarity.

Participant selection into the control group may mean that their results were not representative of the general student population. The fact that the control group contained students from the English speaking society and other co-curricular English learning programs may mean that differences were not observed between the groups because students from the control group performed these other learning activities.

7.11 Conclusion

Changes in student usage around the half-way mark in the study indicate the importance of deploying mobile systems in classroom settings for extended periods of time. The nature of course curricula and student perceptions or motivations that are illustrated through the observed usage patterns means that the majority of studies, which have been conducted over fewer than five weeks, do not capture the full life cycle of these applications within the context of a course or even student perceptions of the usefulness of the MALL approach that is being studied.

This study provides evidence for the effectiveness of using a particular approach to adaptive MALL to support vocabulary learning in formal educational environments. This evidence comes from objective measures of student vocabulary knowledge, subjective measures of their language abilities, and models tying their English knowledge to their system usage (RQ3). This study details some of the ways in which these types of learning-support systems can be integrated into
classrooms to support different activities (RQ2). Unlike the proof-of-concept study and formative evaluation of this approach to MALL, this study showed that the scaffolding provided through VocabNomad can support structured learning activities that reinforce curricular goals as well as the more self-guided learning activities that were observed in Chapter 3 and Chapter 6. The use of this personalized approach to supporting learning activities across educational and cultural contexts indicates that the tool holds promise for helping ELLs with varied backgrounds and needs.

This study provides evidence that continued use of the MALL tool influences learners’ self-perception of their ability to communicate in the target language. However, this was not directly measured and will be more thoroughly addressed in Chapter 8 even though the proof-of-concept study and formative evaluation of VocabNomad provide evidence of the tool’s ability to support ELL communication.
Chapter 8 - Affect and Communication Study in an English Language Environment

The prior studies of VocabNomad use by ELLs have focused on how ELLs have used VocabNomad to support their learning or communication activities (Chapter 6) as well as the relationship between tool use and vocabulary knowledge (Chapter 7). However, these studies have not measured VocabNomad’s relationship to ELL communicative success or affective state, both of which are argued to influence language learning. Since the affective filter or our affective state (Krashen, 1982), our interactions with language (Graves, 2013; Wagner et al., 2007), and our use of language (Swain, 1995) can influence our learning, this study explores the ability of English language learners to achieve their communication goals (i.e., the user’s perceived communicative success). This study also explores the relationship between user affect and MALL tool use because of affect’s influence on learning.

Following from theories of how affect and language production influence language learning, we explored two research questions.

**RQ6:** What is the relationship between the use of a mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary usage and ELL communicative success when that tool is used in an English-language dominant environment?

**RQ7:** How does ELL affective experience change when they have access to a mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary usage?

These questions can be answered both qualitatively and quantitatively and contribute towards answering the primary research questions of VocabNomad’s influence on ELL communication and affect (RQ1 and RQ3) as well as how ELLs integrate such systems into their learning and communication practices (RQ2). When considering these questions in the light of prior work, four hypotheses emerge. These hypotheses are based on operationalizations of an ELL’s experience and focus on the learner’s affective state and communicative success.
H2: The introduction of a mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary use will be followed by an improvement\textsuperscript{11} in the affective state of ELLs.

H3: Following removal, the learner’s affective state will exceed\textsuperscript{12} baseline levels.

H4: The introduction of a mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary use will be followed by an increase in ELL communicative success.

H5: Communicative success following the tool’s removal will exceed baseline levels.

8.1 Methodology

A single-subject with reversal design was used (Jackson, 2012; Krishef, 1991). This consisted of three phases that were conducted in the following order: baseline, treatment or VocabNomad, and reversal. Each phase lasted a minimum of three days and transitions between phases were based on measures of participant behaviours. To transition from one phase to the next, a participant’s measures were supposed to have stabilized. A minimum of three samples was needed to determine measure stability.

The measures that were tracked to determine when to change study phases were collected using the experience sampling methodology (ESM) (Hektner et al., 2007). ESM is a self-report based method that collects data about a person’s everyday experiences. Using experience sampling allowed us to capture detailed, in-situ information about participant experiences. Given the goals of this study, the self-report data that was collected using experience sampling focused on the participant’s affective state and communication experiences.

\textsuperscript{11} An improvement can be shown through a reduction in negative affect or an increase in positive affect.

\textsuperscript{12} The negative affect score will be lower or the positive affect score will be higher than those obtained during baseline measurements.
Experience sampling collects rich data that allows us to characterize ELL experiences. It also enables the frequent and regular tracking of the daily affect and communicative success measures that is necessary for enabling a single-subject design (Krishef, 1991). This level and style of tracking allows for the identification of changes in participant behaviour at the individual’s level. Single subject with reversal designs provide a reasonable approach to tracking changes in participant behaviours, especially when resources are limited or it is difficult to obtain access to a large homogeneous sample, but this choice in study design limits generalizability to a population.

8.1.1 Recruitment and Consent

Participants were recruited from an academic English language training program in the greater Toronto area. I visited their course to introduce myself and explain the research project. Prior to performing recruitment, I had met with the course instructor and the program coordinator to explain the study. It was agreed that students could use the devices as they saw fit. Further conversation with the instructor led to vocabulary from the academic word list (Coxhead, 2000) being added to the learning content that participants were given through VocabNomad.

Participation was voluntary and those who completed the study could earn up to 100 dollars based on their participation activities. Participants received 20 dollars for returning the device, 0.50 dollars for each sample that they submitted, and 40 dollars for submitting at least 90 percent of the samples. Remuneration was capped at 100 dollars with participants receiving 71.67 dollars ($D = 14.76$) on average. Additional details about participant remuneration can be seen alongside their demographic information in Table 39.

8.1.2 Data Collection and Measures

The majority of measures and data were collected through the experience sampling application that we built (i.e., MyLog). These data included measures of participant affect and communicative success. Participant demographic information and assessments of language knowledge were conducted in person using paper instruments. Participant interviews were conducted at the same time as the language assessments. These interviews collected qualitative information and allowed for the clarification of information that had been collected through MyLog.
8.1.2.1 Affect

Affect was measured because there are theories and empirical studies that argue its importance within language learning (Dixon et al., 2012; Krashen, 1982; Swain, 2013). Two measures of affect were used since each has its own strengths and weaknesses. The self-assessment manikin (SAM) (Suk, 2006) is a graphical instrument and the International Positive and Negative Affect Scale Short Form (I-PANAS-SF) is language-based (Thompson, 2007).

The five-point portrait version of SAM was used (Suk, 2006). It shows a graphical representation of five affective states along a range between a highly-positive valence expression and a highly-negative valence expression, but it has not been validated for use cross culturally. SAM asks people to select the image that most closely represents their affective state for a given period of time (Figure 35). It was used because it enables the collection of affect data without requiring that participants understand the abstract English vocabulary that is associated with different affective states.

The I-PANAS-SF is an additive scale that has been validated with people from different cultures, none of whom were reported to have limited English-language abilities (Thompson, 2007). It consists of 10 emotional descriptors of varying arousal levels; five are associated with positive valence and five are associated with negative valence. Each set of these descriptors form a subscale: one for positive affect and one for negative affect. The I-PANAS-SF asks people to

![Figure 35 The affect measures as they appear in the experience sampling application: SAM with momentary instructions (left; black) and the I-PANAS-SF (all others). The momentary (centre-left; black), daily reflective (centre-right; yellow), and weekly reflective (right; green) time instructions are shown for the I-PANAS-SF.](image)
rate how much they feel each of the descriptors for a specified period of time. I-PANAS-SF ratings were performed using the standard 5-point Likert rating scale (see Figure 35), but the administration method was modified slightly. Participants were asked to rate their responses on the scale using radio buttons rather than writing numbers beside the descriptors. This made it easier to ensure that appropriate data were being entered, it made responding faster for participants, and it allowed all of the items to be displayed on the same screen.

Both affect measures were administered with three different time instructions (Figure 35). The measures of the participant’s affect for the current moment were administered three times per day. The reflective self-report of the learner’s affect over the course of the day was collected at the end of the day, and a weekly reflective report on the participant’s affective state over the course of the previous seven days was collected at the end of each week. If a study phase lasted more than seven days, the weekly affect measures were re-administered at the end of the phase.

8.1.2.2 Communication: Number of Attempts and Perceived Success Rates

Participants were asked to report on the number of times that they attempted to communicate in English, what they were most recently trying to communicate, and their perceived level of success during that attempt at communication using English. These measures are linked to theories that argue the importance of language production to language learning (Swain, 1995, 2006)

![Figure 36](image)

Figure 36 The communication measures as they appear in the experience sampling application: both the momentary (left; blue) and the reflective daily (right; yellow) versions are shown.
Two versions of this measure were collected. The first is a momentary self-assessment that was administered at two different times during the day. The second was administered at the end of the day and asked the learner to reflect over all of the communication that he or she had attempted on that day. Figure 36 shows the details of the forms that participants completed.

8.1.2.3 Perceived Competence Scale (PCS)

The perceived competence scale measures a person’s belief in his or her ability to do something (Deci & Ryan, 2002), which has theoretical ties to learning (Dixon et al., 2012; Swain, 2013). For the purposes of this study, the PCS was modified to measure two things: a person’s confidence in his or her ability to learn English and a person’s confidence in his or her ability to communicate using English. Both of these were administered at the end of each study phase. All scale items are rated on a 7-point Likert-type scale, where 1 indicates “not at all true”, 4 indicates “somewhat true”, and 7 indicates “very true” (Figure 37).

8.1.2.4 MyLog – An Experience Sampling Application

All of the communication and affect measures were collected via an experience sampling application called MyLog. To ensure that participants were signaled when they were awake, the application required users to set the times at which they normally wake up and go to sleep. The period between the wake and sleep times was sub-divided into 6 sampling periods, and participants were sampled or signaled at a random time within each sampling period. This sub-
division and randomization of sampling helped ensure that participants were not always signaled while performing the same task. It would have been difficult to characterize participant experiences if they were always asked about their affective state or communicative success during their morning workout.

The first, third, and fifth sampling periods collected momentary affect data. The second and fourth sampling periods collected momentary data about participant communication. The sixth sampling period collected data about the participant’s day as a whole and included the participant’s reflective rating of his or her affect and communication over the course of the entire day. At the end of a study phase or every 7 days within a study phase, participants were asked to report on their communication and affect for the previous week. This is when they were asked to complete the PCS, which indirectly measures their communicative abilities.

Participants could skip any momentary signal (Figure 38) by postponing it past the end of the current sampling period. If a participant chose to postpone completing a sample, he or she was re-signaled 30 minutes later provided the participant was still within the same sampling period. So, if my current sampling period is scheduled to end at 11am and I choose to skip the signal that I receive at 10:15am then I will be re-signaled at 10:45am. If I again postpone the survey, I will not be allowed to complete that sample because the next sampling period will have started. Daily reports could not be skipped. They could only be postponed. This was done to ensure that the

![Figure 38 MyLog: application signaling and postponing interface (left; black) and time instructions (all others): momentary affect (left-centre; black), daily (right-centre; yellow), and weekly (right; green).]
measures that were being used for transitioning between study phases were collected.

The Action Bar colour and first screen of the application were modified to communicate the time instructions (Figure 38). The Action Bar colour also communicated the type of survey that was being conducted. Participants were always asked to provide contextual information that included data about who they were with and what they were doing (Figure 39). For open-ended responses, participants could type an answer, take a photograph, or record an audio response. If they chose to type their response, the android Auto Complete feature could be used to help them compose their message. This also helped to prevent limited participant spelling knowledge from affecting the quality of the data that was being collected.

Regardless of the time instructions that were given during sampling, all of the affect screens collect the data shown in Figure 35. The only thing that changes is the time designation within the instructions. The time instructions can have one of three values: right now, today, and during the last week. The affect surveys also ask participants to explain the reason behind their current affective state and any new activities that they have performed on that day (Figure 39).

Participants were trained in the use of the application during our first meeting, where they were asked to complete one of each of the surveys. The signaling and postponing of surveys was also performed during this session.
Table 38 The data collection schedule for the communication and affect study. Black circles indicate when a measure was taken in relation to the study stages.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Baseline</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Post-Baseline</td>
<td></td>
<td></td>
<td>•</td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-VocabNomad</td>
<td></td>
<td></td>
<td>•</td>
<td>•</td>
<td></td>
<td></td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>Post-Reversal</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td></td>
<td></td>
<td>•</td>
<td></td>
</tr>
</tbody>
</table>


8.1.2.5 Demographic Information

The demographics form used to collect information about the learner’s background was similar to that used in the vocabulary acquisition study. It collected information about the learner’s mother tongue and exposure to English, family member knowledge and use of English, the amount of time that the learner has spent in English-language environments, the English-language media that they have used, and the other learning activities that they have performed or were currently performing. The demographics form also asked participants to rate their proficiency in English and any other languages that they know. The demographics form can be seen in Appendix B. This form was administered in the presence of the primary researcher during my second meeting with participants.

8.1.2.6 Interviews

Semi-structured interviews were conducted in person. The protocol for these interviews was similar to those used in the formative evaluations of VocabNomad. Interviews focused on language learning and English communication as well as participants’ experiences with technology and the provided application. The interview guides that were used can be seen in Appendix C. The final interview was also used to seek clarification of data that was submitted through the ESM application.

8.1.2.7 Language Knowledge

The language proficiency measures were administered in-person. Their administration schedule can be seen in Table 38. The majority of these measures are the same as those administered during the vocabulary learning study that was conducted in Japan (Chapter 7). They are not re-
explained here. Their definitions and a description of their administration methods can be seen in 7.1.4.4 Assessments of Language Knowledge.

The measures include learner self-reports of their knowledge, two measures of morphological knowledge (i.e., Morpho-Orthographic Choice Task and Carlisle’s Morphological Structure Test), and one measure of receptive vocabulary knowledge (i.e., PPVT-4). VocabNomad’s adaptive testing feature was used to create adaptive and personalized tests (Person. Test in Table 38). However, these tests were administered in person and scored by hand. Additional details about these tests and their scoring can be seen in 5.3.3.3 Adaptive Theory and Appendix H.

8.1.3 Participants

Twelve ELLs participated in the study: 6 female and 6 male (see Table 39). Their average age was 23.23 years ($SD = 4.61$). All participants were registered in an academic English training program and recruitment had been performed through one of their courses. The majority of participants were Brazilian speakers of Portuguese. Five had knowledge of languages other than English and their mother tongue(s), and five participants used more than one language in their homes (Pluri. Home column from Table 39).

<table>
<thead>
<tr>
<th>ELL</th>
<th>Age</th>
<th>Sex</th>
<th>Mother Tongue</th>
<th>Pluri.</th>
<th>Mos. in Canada</th>
<th>Other Languages</th>
<th>Completed Samples (%)</th>
<th>Days in Study</th>
<th>Remuneration ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESM01</td>
<td>19</td>
<td>M</td>
<td>Portuguese</td>
<td>✓</td>
<td>3</td>
<td>Spanish</td>
<td>83.33</td>
<td>26</td>
<td>85.00</td>
</tr>
<tr>
<td>ESM02</td>
<td>20</td>
<td>M</td>
<td>Portuguese</td>
<td></td>
<td>3</td>
<td>78.47</td>
<td>24</td>
<td>76.50</td>
<td></td>
</tr>
<tr>
<td>ESM03</td>
<td>23</td>
<td>M</td>
<td>Portuguese</td>
<td></td>
<td>3</td>
<td>58.67</td>
<td>25</td>
<td>64.00</td>
<td></td>
</tr>
<tr>
<td>ESM04</td>
<td>20</td>
<td>M</td>
<td>Portuguese</td>
<td></td>
<td>3</td>
<td>62.67</td>
<td>25</td>
<td>67.00</td>
<td></td>
</tr>
<tr>
<td>ESM05</td>
<td>21</td>
<td>F</td>
<td>Portuguese</td>
<td>✓</td>
<td>3</td>
<td>81.33</td>
<td>25</td>
<td>81.00</td>
<td></td>
</tr>
<tr>
<td>ESM06</td>
<td>23</td>
<td>F</td>
<td>Portuguese</td>
<td>✓</td>
<td>3</td>
<td>60.67</td>
<td>25</td>
<td>65.50</td>
<td></td>
</tr>
<tr>
<td>ESM07</td>
<td>24</td>
<td>F</td>
<td>Portuguese</td>
<td></td>
<td>2</td>
<td>Spanish, French</td>
<td>85.33</td>
<td>25</td>
<td>84.00</td>
</tr>
<tr>
<td>ESM08</td>
<td>21</td>
<td>F</td>
<td>Portuguese</td>
<td></td>
<td>3</td>
<td>Japanese</td>
<td>74.73</td>
<td>31</td>
<td>89.50</td>
</tr>
<tr>
<td>ESM09</td>
<td>23</td>
<td>F</td>
<td>Kazakh</td>
<td>✓</td>
<td>9</td>
<td>Russian, Korean</td>
<td>81.94</td>
<td>24</td>
<td>79.00</td>
</tr>
<tr>
<td>ESM10</td>
<td>37</td>
<td>M</td>
<td>Arabic</td>
<td>✓</td>
<td>10</td>
<td>56.17</td>
<td>27</td>
<td>65.50</td>
<td></td>
</tr>
<tr>
<td>ESM11</td>
<td>20</td>
<td>M</td>
<td>Portuguese</td>
<td></td>
<td>3</td>
<td>81.67</td>
<td>20</td>
<td>69.00</td>
<td></td>
</tr>
<tr>
<td>ESM13</td>
<td>23</td>
<td>F</td>
<td>Uygur/Kazakh</td>
<td></td>
<td>3</td>
<td>Russian</td>
<td>14.14</td>
<td>33</td>
<td>34.00</td>
</tr>
</tbody>
</table>

Pluri. = Plurilingual
In addition to showing basic demographic information, Table 39 shows how much each participant received in exchange for his or her participation as well as his or her compliance with the study protocol. This is represented as the percentage of samples that each participant completed. On average participants submitted 68.26 percent ($SD = 20.13$) of the questionnaires for which they were signaled. ESM13’s compliance was too low (14 percent), so her MyLog collected self-report data was excluded from later analyses.

8.2 Results

I first describe how participants used different technologies to support their language learning and communication.\(^{13}\) I then detail participant uses of VocabNomad and MyLog before presenting evidence that user affect and communicative success improved when using the supplied MALL tool.

8.2.1 Technology Use

The description of participant technology use for supporting their English-language learning that is provided below was based on the responses that participants provided through the demographics form and their interviews (Figure 40). Overall, participants were aware of their language-learning processes and the types of challenges that were typically faced by speakers of their mother tongues. They were also aware of the limitations of the technologies that they were using.

Even though participants aligned themselves with interactional approaches to language learning, they primarily used technology to fill the gaps in their knowledge that they had noticed. With the exception of ESM01, ESM07, and ESM11, participant reports did not indicate that these gaps were the result of interacting with others in English. Their secondary use of technology was to increase their exposure to English. This is where all participants reported experiencing gaps in their knowledge that they filled by relying on a combination of people and technologies.

All participants used a variety of technology-based English-language media (Figure 40). This includes movies, television, and music. Media use, such as books and articles, is not included

\(^{13}\) Initial findings appeared in Mboutsiadis and Demmans Epp (2015). See Appendix L for a complete list of publications and their relationship to my thesis work.
Figure 40 Advanced ELL technology use to support language-learning activities.

here since they are not the type of technology that is the focus of this work. When participants used sub-titles to watch video-based media they would use their L1 and L2 in combination but not necessarily in the way that others might expect. For example, ESM05 would listen to the English audio but use Portuguese sub-titles for reference and support.

Participants used a combination of translators and dictionaries based on their need for accuracy and the time constraints that they were facing. Interestingly, the Portuguese-speaking participants expressed concern over the accuracy of Google Translate and would often seek additional information to confirm the translation that had been given. Much of this concern arose around the translation of false cognates, expressions, or phrases that might be colloquial. Some participants would refer to urban dictionary when they wanted more information about a phrase or expression. When participants needed less precision, they would sometimes use Google image searches to obtain a general sense of a word’s meaning and ESM01 used regular searches to obtain usage examples for words since he found this more useful than translation.
All participants, except for ESM10, used some form of mobile application or mobile-accessible services through their mobile devices. Most often, these were installed dictionary applications and the Google Translate application or website. Many participants reported using the dictionary.com application that was encouraged by their instructor, but they reported being frustrated by the lack of available words within the application because it required the download of word packs.

Participants used two types of games to support their language development: language-learning games and online multi-player games. They played multi-player games with people from other countries. In this case, their text-based chat was conducted in English and they often listened to other players speak in English. However, participants struggled to produce oral language in these settings because of the fast-paced nature of the game play and conversation.

8.2.2 VocabNomad Use

Like with the other language-learning technologies, participants used VocabNomad to fill gaps in their knowledge. This could be because they did not understand a word or phrase from a show that they were watching (ESM01) or because they did not know the word that they wanted to use when writing an essay (ESM05).

All participants used VocabNomad to perform some type of direct study by self-testing, reviewing words, or building their vocabularies through the exploration of the adaptively-recommended synonyms (Figure 41).

Only one participant used the system to support her communication in the form of an essay that she was writing for a class. No one used the application to directly support his or her oral communication even though some participants used it to develop their speaking and listening skills (Figure 41).

In addition to the above information about participant activities and the intent behind their actions within VocabNomad, we have the usage logs that are the source of the descriptive statistics in Table 40. These statistics show reasonably high usage levels and are reported in quartiles since the observed event distribution was often non-normal. Most of the events were
logged through the mobile device. However, all participants edited some vocabulary through the web interface and eight ELLs imported collections of vocabulary through the web interface.

Only ESM10 dismissed any of the synonyms that were recommended when he was using the mobile application. Participant information seeking activities were performed through the application with no L1 searches being recorded and most search activities being dedicated to finding specific words rather than categories of words. The majority of participant editing activities involved adding tags to existing vocabulary even though participants did not typically use the tags to search for vocabulary.

Participants did not record themselves or anyone else pronouncing any of the learning content but occasionally listened to the text-to-speech engine verbalize individual vocabulary entries.

8.2.3 MyLog Use

Participants reported that they sometimes started to answer a survey and then would get interrupted or distracted; this meant that the sample was not submitted or was completed at two different times. Participants reported that they sometimes missed signals when they did not
Table 40 Selected VocabNomad usage statistics (event counts).

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Min</th>
<th>Q1</th>
<th>Mdn</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edited a definition (web)</td>
<td>1</td>
<td>10</td>
<td>23</td>
<td>26</td>
<td>41</td>
</tr>
<tr>
<td>Browsed vocabulary (web)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Imported vocabulary (web)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>English word search (app)</td>
<td>240</td>
<td>399</td>
<td>532</td>
<td>676</td>
<td>1240</td>
</tr>
<tr>
<td>Tag search (app)</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td>Added vocabulary (app)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Viewed detailed vocabulary entry (app)</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Added tags (app)</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Unshared vocabulary (app)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Requested a definition for a word (app)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Browsed vocabulary (app)</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>43</td>
</tr>
<tr>
<td>Synonym recommended (app)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>Viewed synonym (app)</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Listened to a vocabulary entry (app)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>All Events</td>
<td>255</td>
<td>453</td>
<td>559</td>
<td>689</td>
<td>1313</td>
</tr>
</tbody>
</table>

notice the phone vibrating because they were too focused, were in the shower, or had forgotten their phone somewhere else. This happened with high and low compliance participants.

Participants admitted to ignoring MyLog signals when they were focused on another task or felt that they were too busy to complete the survey. They also reported ignoring the signal or being selective with respect to the media that they used to respond based on their current social context. For example, ESM11 reported not answering questions when he was in the bathroom at school and ESM06 did not take photographs when in the theatre. However, the rules that participants employed to determine when it was appropriate to answer surveys varied from person to person: ESM06 completed surveys in all social situations and explained what he was doing to those who he was with and ESM09 did not answer surveys when her hands were dirty because she was cooking.

Participants discussed how the MyLog application influenced them. ESM11 felt that it became part of his natural process. ESM01 and ESM13 felt that it increased their awareness of the ratio with which they used their L1 and L2. In ESM13’s case, she reported increasing her English usage as a result of the MyLog application: if she had not used English with someone else when she received the communication-focused survey, she would call someone to make sure that she
had used English that day. Several participants also reported that it was difficult to describe why they felt certain emotions.

Even with participants failing to notice when they were signaled or choosing not to respond as a result of their current context, sufficient data was collected to enable the analysis of the communicative success and affect of most participants from a single-subject design perspective.

8.2.4 Affect

I first present the improvements in participants’ momentary affect. This is followed by a discussion of the observed improvements in participant daily affect.

8.2.4.1 Momentary Affect

The momentary I-PANAS-SF was sufficiently reliable ($\alpha = .76$) and participant responses for the momentary SAM and I-PANAS-SF affect measures were moderately correlated ($r(563) = .48, p < .001$). These statistics show that the measures of affect that were used were evaluating similar constructs.

Descriptive statistics for the affect measures are presented by study phase in Table 41. Participant responses for momentary measures of affect, showed that the positive subscale from the I-PANAS-SF was strongly correlated between the baseline and VocabNomad ($r(9) = .88, p < .001$), the VocabNomad and reversal ($r(9) = .78, p < .01$), and the baseline and reversal phases ($r(9) = .72, p = .01$). Like the positive subscale, the negative subscale of the I-PANAS-SF was strongly correlated between the baseline and VocabNomad phases ($r(9) = .73, p = .01$) but not the VocabNomad and reversal phases ($p = .15$).

Momentary average positive affect was not normally distributed as measured by the Shapiro-

<table>
<thead>
<tr>
<th>Table 41 Momentary affect scores (%) by phase.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>I-PANAS-SF</td>
</tr>
<tr>
<td>Positive (M)</td>
</tr>
<tr>
<td>Positive (Min.)</td>
</tr>
<tr>
<td>Negative (M)</td>
</tr>
<tr>
<td>SAM</td>
</tr>
</tbody>
</table>
Wilk test \( (p < .001) \). A Friedman ANOVA by ranks showed a significant difference in participant average momentary positive affect \( (\chi^2(2) = 6.55, p = .04) \) based on study phase. A related-samples Wilcoxon signed rank test indicated that participant positive affect was significantly higher \( (Z = -2.22, p = .03) \) during the VocabNomad phase than the baseline phase. There was no significant difference between the baseline and reversal \( (p = .21) \) or reversal and VocabNomad phases \( (p = .42) \), which indicates that participant affect was better when they had access to VocabNomad.

Participant minimum positive momentary affect also varied significantly by study phase \( (\chi^2(2) = 6.40, p = .04) \), and a related-samples Wilcoxon signed rank test showed significant increases between the baseline and VocabNomad phases \( (Z = -1.97, p = .05) \) and the baseline and reversal phases \( (Z = -2.41, p = .02) \) but not the VocabNomad and reversal phases \( (Z = -.72, p = .47) \). The results of these tests show an improvement in minimum positive affect that was maintained.

No significant differences were found between phases based on participant maximum positive momentary affect \( (p = .21) \). Participant average \( (p = .91) \), minimum \( (p = .17) \), and maximum \( (p = .05) \) scores on the momentary I-PANAS-SF showed no significant changes based on study phase. The same is true for participant responses on the momentary negative affect subscale from I-PANAS-SF \( (p = .23) \) and the SAM \( (p = .98) \).

8.2.4.1 Daily Affect

Independent of study phase, participants’ daily SAM and I-PANAS-SF (Table 42) were significantly correlated \( (r(216) = .51, p < .001) \). As were the daily SAM responses \( (r(9) = .64, p = .03) \) and the daily I-PANAS-SF responses \( (r(9) = .73, p = .01) \) from the baseline and VocabNomad phases. These correlations indicate that these scales were measuring similar constructs.

<table>
<thead>
<tr>
<th>Table 42 Daily affect scores (%) by phase.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>SAM</td>
</tr>
<tr>
<td>I-PANAS-SF</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
</tr>
</tbody>
</table>

175
Daily participant negative affect was correlated between the VocabNomad and reversal phases \((r(9) = .75, p < .01)\) and participants’ daily positive affect was strongly correlated between the baseline and VocabNomad \((r(9) = .76, p < .01)\), baseline and reversal \((r(9) = .71, p = .02)\), and VocabNomad and reversal \((r(9) = .94, p < .001)\) phases.

Participant daily responses to the SAM \((p = .16)\), positive subscale \((p = .59)\), negative subscale \((p = .27)\), and I-PANAS-SF \((p = .54)\) showed no significant difference based on study phase when testing group statistics.

However, individual participant data shows differences for some affect measures based on study phase. These differences will be explained through the visual analysis of data from three participants. We can see from Figure 42 that ESM11 experienced both a slope and level change in his I-PANAS-SF score, but he did not have enough daily samples in the reversal phase to see how the removal of VocabNomad might relate to participant affect. The effect size of this change between the baseline and VocabNomad phases can be represented using the improvement rate difference (IRD), where values above 50 percent are better than chance (Parker, Vannest, & Brown, 2009; Wendt, 2009). ESM11’s IRD for I-PANAS-SF is 50.91 percent. If we look at ESM11’s response patterns in more detail by decomposing the I-PANAS-SF into its subscales, we can see that the slope change is a result of a change in the direction of the participant’s negative affect, which is an improvement, and the level change is a result of a level increase in the participants’ positive affect.

If we look at the negative-affect subscale from the I-PANAS-SF for ESM02, we see a change in his negative affect with both the level and slope improving (i.e., they are decreasing) during the VocabNomad phase \((IRD = 85.71)\) and increasing (i.e., degrading) upon reversal \((IRD = -71.43)\). ESM04 shows stabilization in his negative affect measures following the introduction of VocabNomad and a similar increase in negative affect from the VocabNomad to the reversal phase \((IRD = -80.00)\).

8.2.4.2 Relationship Between Affect and Application Usage

Participants with higher baseline levels for maximum momentary negative affect had lower application usage levels \((r(9) = -.65, p = .03)\). There was also a moderate positive relationship between system usage and the minimum negative momentary affect that was reported by
participants ($r(9) = .61, p = .05$): participant negative affect was greater during the reversal phase than it was during the VocabNomad phase for those who used the system more extensively.

The weekly SAM scores of participants who used MyLog more heavily were lower ($r(9) = -.76, p = .05$). Similar relationships existed between MyLog usage and the minimum momentary I-

---

**Figure 42** Participant tracking information for daily affect measures; dashed lines indicate the boundaries between phases.
PANAS-SF scores \( (r(9) = -0.70, p = 0.02) \) as well as MyLog usage and the minimum momentary positive affect score \( (r(9) = -0.60, p = 0.05) \) of participants.

### 8.2.5 Communication: Number of Attempts and Perceived Success Rates

Improvements in participants’ perceptions of their communicative success for individual attempts at communication are presented. This is followed by a discussion of improvements in their communicative success over the course of each day. Changes in the amount of communication attempted by participants are then presented as are the relationships between application usage and participant communication.

#### 8.2.5.1 Perceived Momentary Communicative Success

Momentary reports of participant communicative success (Table 43) were moderately correlated between the VocabNomad and reversal phases \( (r(9) = 0.61, p = 0.05) \), with participant average momentary communicative success varying significantly by study phase \( (\chi^2(2) = 7.82, p = 0.02) \). A Friedman ANOVA by ranks was used because the VocabNomad phase data was not normally distributed \( (p = 0.01) \). Wilcoxon signed ranks tests showed significant improvement in participant momentary communicative success between the baseline and VocabNomad \( (Z = -1.867, p = 0.06) \) and the baseline and reversal \( (Z = -0.26, p = 0.01) \) phases. No significant difference was found between the VocabNomad and reversal phases \( (p = 0.53) \).

Participant responses for their minimum momentary communicative success were significantly different between phases \( (\chi^2(2) = 6.82, p = 0.03) \) as measured by a Friedman ANOVA by ranks test: the distribution of participant responses was not normal \( (p = 0.01) \) for the VocabNomad phase. A Wilcoxon signed ranks test showed a significant increase between the baseline and VocabNomad phase \( (Z = -2.09, p = 0.04) \).

<table>
<thead>
<tr>
<th>Table 43 Momentary communication success scores (%) by phase.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>( M )</td>
</tr>
<tr>
<td>Success ( (M) )</td>
</tr>
<tr>
<td>Success ( (Min.) )</td>
</tr>
<tr>
<td>Success ( (Max.) )</td>
</tr>
</tbody>
</table>
Table 44 Daily communication success scores (%) by phase.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th></th>
<th></th>
<th>VocabNomad</th>
<th></th>
<th></th>
<th>Reversal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>79.20 (17.65)</td>
<td>35.33</td>
<td>99.00</td>
<td>82.98 (19.80)</td>
<td>36.00</td>
<td>97.84</td>
<td>87.72 (11.09)</td>
<td>66.50</td>
</tr>
</tbody>
</table>

From the participants’ perspective, no differences were found in the self-report data for their maximum momentary communicative success ($p = .82$).

8.2.5.2 Perceived Daily Communicative Success

Participants’ daily communicative success (Table 44) was moderately to strongly correlated between the baseline and VocabNomad ($r(9) = .61, p = .03$), VocabNomad and reversal ($r(9) = .74, p < .01$), and baseline and reversal ($r(9) = .60, p = .04$) phases.

Participant reports of their daily communicative success were not normally distributed ($p = .001$) and a Friedman ANOVA by ranks showed a significant difference in how participants perceived their communicative success based on phase ($\chi^2(2) = 8.17, p = .02$). A Wilcoxon signed ranks test showed no significant difference between participant success during VocabNomad and reversal ($p = .53$). However, a significant increase was observed between the baseline and reversal ($Z = -3.06, p < .01$), and an improvement approaching significance was observed between the baseline and VocabNomad phases ($Z = -1.65, p = .10$).

The group level statistics are supported by the graphical analysis of the single-subject design tracking for some participants (Figure 43). The stability in ESM07’s communicative success near the end of the baseline phase allows us to compare the relatively flat slope of that line to the somewhat positive slope that exists in the VocabNomad phase (IRD = -25.00), which shows no real difference, and the negative slope that is observed upon reversal (IRD = -71.43). While the measure does not demonstrate a level change, the change in slope between the VocabNomad and reversal phases indicates that ESM07 experienced more communicative success when in the VocabNomad phase of the study. The high variability of ESM05’s communicative success during the baseline phase does not allow us to use that period for comparison. However, the measures are more stable in subsequent phases which allows for their analysis. ESM05 demonstrated more consistent communicative success during the VocabNomad phase. Her
communicative success level during the reversal phase remained similar to that observed in the VocabNomad phase (IRD = 42.86), and it stabilized at ceiling levels following reversal. Like ESM05, ESM04’s communicative success was more consistent in the VocabNomad phase than it was in the baseline phase. ESM04’s communicative success also decreased during the reversal phase (IRD = - 80.00). Like ESM04 and ESM05, ESM06 shows increased stability in her communicative success during the VocabNomad phase but the instability in the baseline phase prevents any improvements from being measured.

Figure 43 Participant tracking information for daily communicative success; dashed lines indicate the boundaries between phases.

8.2.5.3 Number of Communication Attempts

Beyond learner communicative success Table 45 shows that the amount of communication attempted by participants did not vary significantly between phases ($p = .60$). The number of attempts at communication that were made was moderately correlated between the baseline and
Table 45 Daily communication attempts (number) by phase.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>VocabNomad</th>
<th>Reversal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attempts</td>
<td>11.10</td>
<td>(6.20)</td>
<td>2.20</td>
</tr>
</tbody>
</table>

VocabNomad phases ($r(9) = .62, p = .03$) but not the VocabNomad and reversal phases ($p = .08$) or the baseline and reversal phases ($p = .06$).

8.2.5.4 Relationship Between Communication and Application Usage

To further assess the relationship between communication and VocabNomad usage, correlational analysis was performed. Participant application usage was correlated with the minimum levels of communicative success that were reported using the momentary survey during the reversal phase ($r(9) = .61, p = .05$). Like VocabNomad usage, MyLog usage showed moderate to strong relationships with participant communicative success during the baseline ($r(9) = .69, p = .01$), VocabNomad ($r(9) = .79, p < .01$), and reversal ($r(9) = .66, p = .02$) phases.

This analysis further revealed a moderate relationship between VocabNomad usage and the number of communication attempts that participants made each day during the reversal phase ($r(9) = .62, p = .03$).

8.2.6 Perceived Competence Scale (PCS)

The learning ($\alpha = .916$) and communication ($\alpha = .922$) versions of the PCS were highly reliable (Table 46).

The first administration of the learning version of the PCS was the only one that was not normally distributed ($p = .001$). A Friedman ANOVA by ranks showed no significant difference in participants’ perceived confidence of their ability to learn English ($p = .22$) based on study phase. A repeated measures ANOVA with sphericity assumed ($W = .88, p = .94$), showed no

Table 46 PCS scores (%) by administration.

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th></th>
<th>2nd</th>
<th></th>
<th>3rd</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>PCS – Learning English</td>
<td>83.12</td>
<td>15.08</td>
<td>94.05</td>
<td>7.02</td>
<td>90.82</td>
<td>11.64</td>
</tr>
<tr>
<td>PCS – Communicating in English</td>
<td>78.25</td>
<td>18.38</td>
<td>88.10</td>
<td>4.33</td>
<td>87.76</td>
<td>11.82</td>
</tr>
</tbody>
</table>
Table 47 Participant scores (%) on morphological measures.

<table>
<thead>
<tr>
<th>Task</th>
<th>M</th>
<th>(SD)</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morpho-Orthographic Choice</td>
<td>91.95</td>
<td>(7.10)</td>
<td>79.31</td>
<td>100</td>
</tr>
<tr>
<td>Morphological Structure Test</td>
<td>65.06</td>
<td>(19.09)</td>
<td>38.46</td>
<td>100</td>
</tr>
</tbody>
</table>

statistically significant difference in participant confidence in their ability to communicate using English \((p = .61)\). Participant responses on the PCS (Table 46) indicate that they started with high levels of self-efficacy and had little room to grow, which may explain why no differences were found.

8.2.7 Language Knowledge

8.2.7.1 Morphological Knowledge

Item analysis was performed to improve the reliability of both measures. This resulted in 13 items from the morphological structure test being retained, an \(\alpha\) of .73, and a split-half reliability of .76. Item analysis resulted in 29 of the morpho-orthographic choice task items being kept, an \(\alpha\) of .67, and a split-half reliability of .69. It appears that student performance on the measures is not correlated as expected \((p = .88)\). This might be explained by the low-variability and ceiling effects that were observed on the morpho-orthographic choice task.

Participant scores (Table 47) were normally distributed on the morphological structure test \((p = .55)\) and the morpho-orthographic choice task \((p = .11)\) as measured by the Shapiro-Wilk test.

8.2.7.2 Vocabulary Knowledge (PPVT-4)

Item analysis was conducted to increase the reliability of the tests. This resulted in 64 items from the first administration, 67 items from the second administration, and 69 items from the third administration being used to calculate student scores. It resulted in reasonable levels of test form reliability (see Table 48). The alternate forms reliability is also reasonably high: all the test forms are significantly correlated with one another and the relationships are strong. The first

Table 48 PPVT-4 reliability statistics by administration.

<table>
<thead>
<tr>
<th>Reliability Measure</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach’s (\alpha)</td>
<td>.85</td>
<td>.76</td>
<td>.87</td>
</tr>
<tr>
<td>Guttman Split-half Coefficient</td>
<td>.63</td>
<td>.82</td>
<td>.76</td>
</tr>
</tbody>
</table>
Table 49 PPVT-4 scores (%) by administration.

<table>
<thead>
<tr>
<th></th>
<th>1st M (SD)</th>
<th>Range</th>
<th>2nd M (SD)</th>
<th>Range</th>
<th>3rd M (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>75.91 (10.37)</td>
<td>31.25</td>
<td>83.21 (7.20)</td>
<td>25.37</td>
<td>83.45 (9.86)</td>
<td>37.68</td>
</tr>
</tbody>
</table>

administration was significantly correlated with the second \( r(10) = .71, p < .01 \) and third administrations \( r(10) = .74, p < .01 \). As expected, the morphological structure test was correlated with the first \( r(10) = .59, p = .05 \), second \( r(10) = .70, p = .01 \), and third \( r(10) = .75, p < .01 \) administrations of the PPVT-4.

No outliers were found and student responses on the PPVT-4 (Table 49) were normally distributed for the first \( p = .47 \), second \( p = .73 \), and third \( p = .95 \) administrations as measured by the Shapiro-Wilk test of normality. A repeated measures ANOVA with sphericity assumed \( p = .59 \) showed significant changes in participant PPVT-4 scores based on time of administration \( F(2) = 9.52, p = .001 \), partial \( \eta^2 = .46 \). Post-hoc analysis with Bonferroni correction showed significant improvement between the first and second administrations that defined the baseline phase \( p = .02 \), and the first and third administrations \( p = .01 \). Participant score gains from the baseline phase were significantly higher than those measured during the treatment phase \( t(11) = 2.27, p = .04 \).

8.2.7.3 Vocabulary Knowledge (Adaptive and Personalized Testing)

The Shapiro-Wilk test indicates that student scores (see Table 50) for all of the adaptive and personalized tests were normally distributed (see Table 51). The personalized tests were moderately correlated \( r(10) = .65, p = .02 \) but the results of these tests were not correlated with the results of the adaptive tests that were administered at the same time (Table 52).

Table 50 Shapiro-Wilk test \( p \)-values for the adaptive and personalized testing by administration.

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive</td>
<td>.840</td>
<td>.809</td>
<td>1.00</td>
<td>.320</td>
</tr>
<tr>
<td>Personalized</td>
<td>.939</td>
<td>.259</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 51 Adaptive and personalized testing scores (%) by administration

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>Range</td>
<td>M (SD)</td>
<td>Range</td>
</tr>
<tr>
<td>Adaptive</td>
<td>57.62 (12.29)</td>
<td>44.29</td>
<td>72.02 (10.99)</td>
<td>37.14</td>
</tr>
<tr>
<td>Personalized</td>
<td></td>
<td></td>
<td>49.58 (12.80)</td>
<td>41.25</td>
</tr>
</tbody>
</table>

Table 52 P-values for the relationships between the adaptive and personalized tests.

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>.28</td>
<td>.12</td>
<td>.87</td>
<td>.19</td>
</tr>
<tr>
<td>2nd</td>
<td>.39</td>
<td>.89</td>
<td>.43</td>
<td>.36</td>
</tr>
</tbody>
</table>

A repeated measures ANOVA, with sphericity assumed ($W = .58, p = .39$), showed a significant difference in participant scores on the adaptive test based on time of administration, $F(3) = 5.24$, $p < .01$, partial $\eta^2 = .32$. Post-hoc analysis with Bonferroni correction showed a significant improvement in participant scores during the baseline phase (between administration 1 and 2), $p = .04$, and first and third administrations ($p = .05$).

A Friedman ANOVA by ranks showed a significant difference in participant score gains on the adaptive test based on study phase ($\chi^2(2) = 8.52, p = .01$) and Wilcoxon signed rank tests showed that the baseline score gain was significantly higher than the score gain for the VocabNomad (treatment) phase ($Z = -2.16, p = .03$).

Paired t-tests showed no-significant difference in participant scores on the personalized test ($t(10) = -1.20, p = .257$).

8.2.7.4 Learner Self-Assessments

As expected participant self-assessments of their mother tongue (L1 – see Table 53) were not normally distributed (see Table 54) as measured by Shapiro-Wilk. There were no outliers and the internal reliability for this measure was unexpectedly low: $\alpha$ ranged between .57 and .39 based on time of administration. However, item analysis could not be used to increase reliability since the questionnaire only contains 3 items.
Table 53 Student self-assessed L1 (varied) and English (L2) proficiency.

<table>
<thead>
<tr>
<th></th>
<th>Speak (SD)</th>
<th>Und. (SD)</th>
<th>Read (SD)</th>
<th>L1 (SD)</th>
<th>English</th>
<th>Speak (SD)</th>
<th>Und. (SD)</th>
<th>Read (SD)</th>
<th>L2 (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>9.75 (0.45)</td>
<td>9.92 (0.29)</td>
<td>9.58 (0.52)</td>
<td>9.75 (0.29)</td>
<td>7.21 (0.99)</td>
<td>8.25 (0.97)</td>
<td>8.00 (1.21)</td>
<td>7.82 (0.83)</td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>9.33 (0.65)</td>
<td>9.75 (0.45)</td>
<td>9.33 (0.89)</td>
<td>9.47 (0.50)</td>
<td>7.00 (1.04)</td>
<td>7.67 (0.89)</td>
<td>7.50 (1.24)</td>
<td>7.39 (0.83)</td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>9.42 (0.52)</td>
<td>9.67 (0.49)</td>
<td>9.33 (0.89)</td>
<td>9.47 (0.48)</td>
<td>7.33 (1.16)</td>
<td>7.83 (0.94)</td>
<td>8.00 (1.21)</td>
<td>7.72 (0.96)</td>
<td></td>
</tr>
</tbody>
</table>

Speak – speaking, Und. – understanding, Read – reading

As expected, a Friedman ANOVA by ranks showed no statistically significant difference in student self-assessed L1 proficiency based on time of administration ($\chi^2(2) = 4.80, p = .09$).

Participant self reports of their English language abilities (see Table 53) were normally distributed (see Table 54). No outliers were found and it was reasonably reliable: $\alpha$ ranged between .67 and .84 based on time of administration.

The assumption of sphericity was not violated ($\chi^2(2) = 0.85, p = .45$) so a repeated measures ANOVA was used to test for differences in participant self-assessed English-language abilities. None were found ($F(2) = 2.64, p = .09$, partial $\eta^2 = .47$).

The score gains for English language proficiency were then analysed. The gains observed between the first and second administrations were not normally distributed ($p = .02$) so a Wilcoxon signed ranks test was used. No significant differences were found ($Z = -1.89, p = 0.06$). However, participants who experienced greater gains in their English scores used MyLog less ($r(10) = -.76, p < .01$).

Table 54 Shapiro-Wilk test $p$-values for participant self-assessments of their L1 (varied) and English (L2) abilities.

<table>
<thead>
<tr>
<th></th>
<th>L1</th>
<th></th>
<th></th>
<th>L1</th>
<th></th>
<th></th>
<th>L2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speak</td>
<td>Und.</td>
<td>Read</td>
<td>L1</td>
<td>Speak</td>
<td>Und.</td>
<td>Read</td>
</tr>
<tr>
<td>1st</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.01</td>
<td>.16</td>
<td>&lt;.01</td>
<td>.25</td>
</tr>
<tr>
<td>2nd</td>
<td>&lt;.01</td>
<td>&lt;.001</td>
<td>.001</td>
<td>.04</td>
<td>.03</td>
<td>.16</td>
<td>.23</td>
</tr>
<tr>
<td>3rd</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.001</td>
<td>.09</td>
<td>.07</td>
<td>.12</td>
<td>.25</td>
</tr>
</tbody>
</table>

Speak – speaking, Und. – understanding, Read – reading
8.3 Discussion

8.3.1 Participant Communication (RQ6)

H3 stated that the introduction of a tool, such as VocabNomad, would be accompanied by increases in ELL communicative success. This study supports H3. Participants experienced increases in their minimum levels of communicative success (or worst case measures) from the baseline to VocabNomad phases; they reported a reduction in the amount of failure that they experienced. The additional increase in their average communicative success between the baseline and VocabNomad phases as well as the baseline and reversal phases indicates an improvement. The lack of change in participants’ perceptions of their communicative success in the reversal phase along with ESM05 experiencing ceiling effects indicates that this improvement was maintained following the removal of VocabNomad. This improvement supports H4 and H5 since both the VocabNomad and reversal phases exceeded baseline levels. The direction change in ESM06 and ESM07’s communicative success provides additional support for H3 since the removal of VocabNomad was accompanied by reductions in communicative success. However, the perceived communicative success of these participants during the reversal phase did not exceed baseline levels. Participants experienced similar levels of communicative success, indicating that H4 requires refinement and further study to better understand the conditions under which improvements in participant communicative success are maintained.

These participants may have decided to attempt more difficult communication tasks than they had previously, which could have resulted in lower perceived communicative success during the reversal phase. Another possibility would be that they attempted to communicate more and thus increased their probability of experiencing failure. These explanations are supported by the moderate correlation between communicative success during the VocabNomad phase and the number of times that participants attempted to communicate during the reversal phase. That is, those who experienced success communicating during the VocabNomad phase may have been more willing to attempt communication following the tool’s removal.

The moderate correlation between the number of communication attempts made by participants during the baseline and VocabNomad phases and the strong correlation between the number of
communication attempts during the VocabNomad and reversal phases indicates an increase and maintenance in the communication efforts of participants following the introduction of the support tool. This is supported by the significant increase in the number of communication attempts between the baseline and reversal phases. An increase that approaches significance between the number of communication attempts made during the baseline and VocabNomad phases indicates that VocabNomad use may be related to participants’ willingness to attempt communication in English.

The relationship between MyLog use and participant communicative success indicates that those with better communication skills were more likely to use the application or that it helped develop their communication skills. The direction of influence for this relationship is unknown: it is equally possible that participants with better communication skills were more likely to use the application and that this reinforced and supported the development of their existing communication skills.

The lack of change in participants’ communication-based PCS scores is not surprising given their already high success rates. Moreover, the lack of improvement in participant confidence does not invalidate the improvement in their communicative success as measured by increases in the consistency with which they achieved their communication goals, their increased communicative success, and increases in the amount that participants attempted to communicate using English.

8.3.2 Participant Affect (RQ7)

It was hypothesized that the introduction of VocabNomad would coincide with improvements in ELL affective state (H2). This hypothesis is supported by participants experiencing their highest average momentary positive affect scores when they were in the VocabNomad phase of the study. Their minimum momentary positive affect was also highest during the VocabNomad phase and exceeded baseline levels upon reversal, thus supporting both H2 and H3. Momentary negative affect also appears to have improved over the course of the study for some individuals. However, the negative subscale of the I-PANAS-SF may not have fully captured or sufficiently distinguished between highly negative affective states; ceiling effects rather than a lack of
improvement may have contributed to the absence of observed improvement in some participants’ negative affect.

The correlation between VocabNomad usage and participants’ reversal-phase, minimum momentary negative affect scores indicates that those who used the application more saw their affective state degrade when they were no longer allowed to use VocabNomad. This provides some evidence for H2, but the reversal levels do not appear to exceed baseline levels for negative affect. This implies that future investigations should study negative and positive affect separately since they behave differently: participants’ positive affect scores supported H3 while their negative affect scores did not.

It is worth noting that high negative affect levels during the baseline phase were correlated with low VocabNomad usage levels. While this may be expected, it may have biased the results based on participant receptiveness. The negative correlation between MyLog usage and several measures of affect may indicate that the MyLog application has a negative influence on participant affect as a result of frustration, signaling interruptions, or the burden of completing study protocols. This burden may have moderated any improvements that might have been observed as a result of VocabNomad usage.

8.3.3 Language Knowledge

The increase in PPVT-4 scores from the first administration to the second and third administrations with no significant difference in student scores between the second and third administration shows an increase in student vocabulary knowledge that was maintained. This is supported by tests that compared the score gain from each phase and showed that students learned more between the first and second administrations of the PPVT-4. Similarly, the increase in student scores on the adaptive test occurred during the baseline phase with student scores being maintained through the VocabNomad phase of the studies.

The results of both assessments, when combined, indicate that this learning was not the result of order effects in the PPVT-4. Nor was it the result of their VocabNomad usage. The learning that happened during the baseline phase could be related to their coursework or their use of the MyLog application. Participant reports of the difficulty that they experienced when trying to
express their emotions and the reasons for their current affective state indicate that MyLog resulted in their performing languaging activities, which led to learning (Kukulska-Hulme & Bull, 2009; Swain, 1995). Participants may have learned the language that they needed during the baseline phase of the study. Their not trying new things during subsequent phases may have led to plateaus in participant knowledge because they were not receiving new input or having to produce language that was difficult for them.

The consistency in participant self reports on both the communication and learning version of the PCS may have been influenced by other factors, including the feedback that they continually received in class or increased awareness of their language abilities. This increased awareness could have been the result of reflecting on their communication and reporting their successes and failures through the MyLog application. This interpretation is supported by the negative correlation between MyLog usage and participants’ self-assessed English proficiency. However, the short-term nature of the study may have meant that not enough time had passed to allow for changes in participant confidence in their ability to communicate or learn, especially when participant confidence started at a high level. This may be the case since gains in language knowledge, communication success, and participant willingness to attempt communication were observed. However, the situation is likely more complicated with many of these and other factors interacting.

8.3.4 Measures

The lack of variability in participant responses on the SAM (Table 41) indicates a lack of discriminant validity, especially if SAM’s moderate correlation with the more variable I-PANAS-SF measure is considered.

Participant average positive affect, as represented by the positive subscale of the I-PANAS-SF, was reasonably stable from phase to phase. This is shown through its test-retest reliability, which was higher than any of the other affect measures, and shows a global stability but not the more detailed stability that was expected for measures of positive affect. Participant momentary negative affect was expected to be highly variable, but this was not observed possibly as the result of ceiling effects in some participant’s negative affect measures. Participants may have
been so taxed by their everyday experiences that observing changes in their negative affect was not possible because they had exceeded a personal tolerance threshold.

The high variability in participant responses to the PCS and small sample size may have prevented changes in participant confidence from being detected. PCS scores for both versions were also high during the baseline phase which meant that participants did not have as much opportunity for improvement.

Participants reaching ceiling on the morpho-orthographic choice task reduced this measure’s reliability. However, the measure was still used to assess the validity of various vocabulary measures since there is overlap in their constructs and they were expected to be correlated.

The lack of correlation between the adaptive and personalized tests may be the result of the more simple morphological properties of the words that were being tested. They are included and used because they showed similar results to those obtained using the PPVT-4.

8.3.5 Implications for Language Learning Technology

Given participants’ high negative affect and their inability to see progress when they make small gains, it is important to provide them with additional feedback and encouragement. The MyLog or VocabNomad application could do this through the use of a visualization of the learner’s knowledge, abilities, or affect. These types of visualizations are called open learner models or learning dashboards and can be used to encourage the learner to perform activities that will benefit his or her learning or increase his or her proficiency with respect to using the language (Bull & Kay, 2010; Tsourounis et al., 2014). The use of an open learner model may also help participants realize that their skills are changing and prevent their perceived lack of progress from negatively influencing their adoption of the learning technology (Demmans Epp & McCalla, 2011).

The tracking and journaling that was performed via MyLog seems to have encouraged languaging and noticing, both of which benefit language learning. The definition editing and addition of tags that participants performed also indicate deeper processing activities such as languaging and noticing. Participant use of both applications empirically supports arguments about the potential for MALL to support noticing (Kukulska-Hulme & Bull, 2009). The ability to
perform journaling and recording activities is supported by the results of this study and can be augmented through the use of an open learner model.

Participant frustration over the inability to get all of the words that they wanted from their dictionary because they had failed to predict their needs and preload vocabulary packs indicates that VocabNomad should continue to allow users to import vocabulary. It also suggests that predicting ELL needs and pre-loading vocabulary when the user has a Wi-Fi connection would be beneficial since this would reduce the time required to obtain support when a need arises.

Participant willingness and comfort with using multiple tools to meet their language-support needs indicates that it is okay for VocabNomad to not always meet their needs. However, participants in this study were competent communicators and were proficient in the use of everyday English. The need for the application to provide greater support and better predict user needs may be higher when ELLs are less proficient.

8.4 Limitations

The reliability, order of presentation, or administration method for the various language assessments may have reduced the observability of changes. These factors could have resulted in larger score differences as a result of potential increases in measurement error. The study’s small sample size and design may have contributed to this lack of instrument reliability.

The single-subject design limits the generalizability of the results to the larger ELL population. The study’s short duration along with participant attrition, maturation, and history (i.e., world events) may have influenced study results. Specifically, the short duration may have meant that learning could not be observed. Furthermore, it is likely that interactions between the studied independent variables and other independent variables that were not measured influenced the measures of language knowledge, communicative success, and affect.

8.5 Conclusion

This study expands our understanding of the primary research questions that are the focus of the larger research project. It provides additional information about the extent to which ELLs can use the type of support provided by VocabNomad to scaffold their language-learning and
communication needs (RQ1); how ELLs use this scaffolding to support their vocabulary learning and communication (RQ2); and the relationship between VocabNomad usage and ELL affect, communicative success, and language knowledge (RQ3, RQ6, and RQ7).

Unlike participants from the proof-of-concept study (Chapter 3) and final formative evaluation of VocabNomad (Chapter 6), the ELLs who participated in this study did not use the application to directly support their oral communication (RQ1 and RQ2). This is only surprising when comparing student behaviours from this study against those of the formative evaluation since participants in that study had similar language abilities. So, the choice to use VocabNomad to directly support oral communication does not depend entirely upon ability.

VocabNomad’s support for oral communication in more proficient populations may stem from the indirect support that it provides (RQ1). This support comes in the form of usage examples that can help ensure the subsequent correct use of a word (Godwin-Jones, 2010), which could prevent miscommunication. Additional support for ELL communication may come from the success that participants experienced when they had access to the support that the tool provides (H4) or increased participant confidence and willingness to attempt communication as the result of having access to appropriate scaffolding (RQ3).

While the vocabulary learning-focused study indicated that ELLs showed gains in their vocabulary knowledge when using VocabNomad, this study did not. As a result, there may be limits on the extent to which VocabNomad can support vocabulary learning: it may only benefit learner’s breadth of vocabulary knowledge when ELL vocabulary knowledge is below a particular threshold or they are being exposed to an entirely new vocabulary topic (RQ1).

Participants used VocabNomad to practice their listening skills which contributes to both their language learning and their communication (RQ2). Participants also used the application to support their written communication when completing homework but they did not use it to support their oral communication (RQ2). This decision may depend on various cultural factors, including how people from those cultures manage their reputations (Goffman, 1959) and their relationships to others (Brown & Levinson, 1999). Some cultures are more accepting of mistakes or of asking for help. This allows ELLs from these backgrounds to talk their way through or around communication barriers, whereas those from cultures where mistakes are heavily
penalized or requesting help is viewed as an imposition may not be as comfortable taking the risk of making a mistake which leads them to rely more heavily on support tools when communicating.

Participants used the tool to review and study vocabulary (RQ2). They tried to expand their vocabulary by exploring vocabulary item definitions, using images to obtain a rough meaning for a word, or exploring the synonyms that were adaptively displayed. They also performed self-testing activities and used the application to support their spelling. These activities show substantial integration of the application for the support of basic vocabulary knowledge development.

Unlike the previous summative evaluation of VocabNomad that took place in a classroom setting, there does not appear to be a relationship between vocabulary knowledge and application usage (RQ3). It is possible that the assessments being used in this study were insufficient for fully measuring this relationship since ELLs had higher initial language proficiency and knowledge. It may be that tests that focused more on learner depth of knowledge or their ability to produce language would have shown relationships.

There was a relationship between ELLs affect and VocabNomad usage (RQ3), with improvements in their positive affect observed (H2). VocabNomad use also supported reductions in negative affect for some participants (H2). When combined, this indicates that system usage may support learning through its influence on the learner’s affective filter (Krashen, 1982).

It is worth noting that while some participants expressed similar levels of negative affect to participant OS from the MyVoice study, these ELLs clearly had higher levels of self-efficacy. These feelings of self-efficacy may be partly responsible for their learning and communicative success. Their sense of self-efficacy may also contribute to the increase in the number of communication attempts that participants made after they had used VocabNomad (RQ3).

Increases in participant attempts at communication were observed alongside increased consistency in ELL communicative success after they had been given access to the support tool (H4). In addition to increased consistency, participant communicative success improved
following the introduction of this tool and participants experienced fewer communication failures as measured by their minimum communicative success scores (RQ3).

The above data and study results paint a complicated picture of the role played by the MALL approach that is implemented through VocabNomad. This includes the approach’s influence on the development of ELL communication skills and language knowledge. This picture will be further detailed in Chapter 9, where the results of all studies will be examined in greater detail.
Chapter 9 – Discussion, Future Work, and Conclusion

The design and development of the MALL approach to supporting ELLs that VocabNomad operationalizes followed iterative user-centered practices that were consistent with an evaluation framework for adaptive interactive systems. This design and development included the consideration of language-learning theories and was informed by a variety of theories about language use, learner exposure to language, and the importance of the context in which learners find themselves when performing learning activities. The high-level design of this tool was evaluated using interviews following a short deployment study of a communication support tool. This study aimed to determine whether communication support tools could enable the language learning and communication activities of ELLs (Chapter 3). This study showed that assistive and augmentative communication tools could be used to support ELL activities and resulted in the development of nine design recommendations. Another tool, called VocabNomad, was then developed. VocabNomad employs a new approach to supporting ELL communication and learning activities. Its design addressed each of these recommendations to varying levels and kept the main visual design elements from MyVoice. Café studies were used to further inform the design of the tool’s user interface and ensure its usability; these studies were performed alongside the development of on-demand learning material generation (Chapter 4). The approaches used to generate this learning material were validated through two studies and VocabNomad was validated for its ability to support English language learner activities (Chapter 6). The tool was then deployed in two settings to evaluate how ELLs could use this MALL-based approach to support their learning and communication (Chapter 7 and Chapter 8).

While these evaluations showed that systems like VocabNomad support the learning activities and communication of some ELLs, this support was not universal. The details of which ELL activities were supported in which settings are provided throughout the discussion of the project’s primary research questions. The relationships between tool use and various learner skills or knowledge are also discussed in response to the primary research questions.

Following the discussion of the various study results and how they inform the larger research questions, I present potential expansions on this work. These expansions include opportunities for building on the current system as well as research opportunities that could help develop our
understanding of ELL experiences and ELL usage of different types of support. This includes work that has already begun (Tsourounis & Demmans Epp, in press; Tsourounis et al., 2014).

9.1 Research Questions and Hypotheses

As previously stated, the work explored in this thesis focused on three questions. These questions were informed by existing language-learning theories but did not explicitly aim to further develop these theories. Rather, they seek to confirm the applicability of these theories within this new approach to mobile assisted language learning. The targeted research questions are reviewed here to support the subsequent discussion of the findings that came from the studies that were reported in this thesis.

**RQ1.** To what extent can an adaptive mobile tool support both the vocabulary learning and communication of ELLs?

**RQ2.** How do users integrate an adaptive mobile tool into their language learning and communication practices?

**RQ3.** What is the relationship between the usage of an adaptive mobile tool and ELL vocabulary knowledge, communicative success, and affect?

Individual studies focused on different combinations of these questions. The studies also focused on different aspects of each question, which led to the development of sub-questions and their related hypotheses.

**RQ4.** To what extent can we effectively integrate VocabNomad into an English as a foreign language classroom, when using it as a supplement to other course materials and activities?

**RQ5.** How does VocabNomad use relate to student vocabulary knowledge?

**RQ6:** What is the relationship between the use of a mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary usage and ELL communicative success when that tool is used in an English-language dominant environment?
**RQ7:** How does ELL affective experience change when they have access to a mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary usage?

**H1:** ELL receptive vocabulary will increase following the introduction of an adaptive mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary use.

**H2:** The introduction of a mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary use will be followed by an improvement in the affective state of ELLs.

**H3:** Following removal, the learner’s affective state will exceed baseline levels.

**H4:** The introduction of a mobile tool that provides vocabulary-image pairs that are accompanied by contextually rich examples of vocabulary use will be followed by an increase in ELL communicative success.

**H5:** Communicative success following the tool’s removal will exceed baseline levels.

The remainder of the discussion of the above research questions will be organized by primary research question (RQ1, RQ2, and RQ3). Aspects of the other research questions and hypotheses are integrated into the discussion of the primary research questions.

### 9.1.1 RQ1: Extent of Support for Vocabulary Learning and Communication

VocabNomad supported user communication and vocabulary learning. Not surprisingly, the extent of the provided support is influenced by several factors including the language profiles of users, their cultural background, the broader context in which they are using the application, their comfort with technology, and their willingness to take risks.

We saw increases in public space usage between the initial proof-of-concept study and later studies that were conducted in informal learning settings. This difference in how well
VocabNomad supported learner needs in public spaces may have been due to ELL comfort with using technological supports in public spaces, the tool’s ability to support a wider range of ELL needs than the originally studied tool, or increased learner self-efficacy and proficiency.

Users were able to support their oral communication but only did so when other strategies or support tools had failed them. Their choice to first use translation or other tools when communication failed may indicate that they prize expediency over the support provided by MALL tools. Learner familiarity with other MALL tools and the habits that they had already formed may be another contributing factor. Continued, extended use of VocabNomad may have resulted in changes to learner habits and a greater reliance on the support that it provides.

The creation of primarily cohesive lists of vocabulary items, by the content-generation algorithms, helped the learners obtain and study learning content that was specific to their context. Participant use of the sharing and import mechanisms shows that the scaffolding approach that is employed by this tool supports topic-specific self-directed learning. However, the provided scaffolding did not support emergent learner needs as well as might be hoped. This is partly due to the amount of time that is needed to generate and transfer the learning materials following a request for on-demand support.

High proficiency participants were able to identify noise in the algorithmically generated results (e.g., FormEv8 asked why prenatal was tagged with gym). The ability of learners to identify these types of noise and their tag editing activities indicate that the community curation model can be used to help rectify the inaccuracies that result from using corpora to which anyone can contribute. This type of collaborative, co-construction of knowledge was supported through the editing and sharing of vocabulary. However, participants wanted to directly share content with each other or import content from particular users.

Some users preferentially selected vocabulary entries with images to study and would not study items that did not have images associated with them, whereas other users wanted to study the vocabulary entry’s definition but would still study vocabulary entries that did not have definitions. The preference for images may have been linked to the learner’s existing English vocabulary base with lower-proficiency learners having a stronger preference for visual media. However, some intermediate to advanced learners expressed a preference for images over text.
These learner reports indicate that the MALL approach used within VocabNomad did not fully support vocabulary learning activities when users had specific media preferences that prevented them from studying vocabulary that did not match those preferences. However, it is not always possible or even advisable to support learner preferences since these preferences could be detrimental to learning (Stockwell, 2012): if a visualization might give the learner the wrong idea then it should not be used and other media should be employed. The limiting nature of learner preferences and different media makes mixing media and encouraging learners to experiment with different forms of media and infrequently used features important (Tsourounis & Demmans Epp, in press).

Participants S and R, from the proof-of-concept study, commented on the system not supporting their communication needs when a situation requires a different register (i.e., variety of language based on its social context of use (Crystal, 1991)) than the one with which they are familiar. While participants in subsequent studies did not comment on the support provided for different registers, no improvements were made to explicitly support their use because pragmatics were outside the scope of the MALL approach being explored within this thesis. Moreover, this type of challenge is commonly encountered by immigrants and can be overcome by exposure to different forms of English and instruction on the appropriate use of these forms (Ruhlemann, 2008).

Like its support for various registers, the VocabNomad approach to MALL falls short of providing full rehearsal support in the way that a communication partner would. This means that users who wish to practice a conversation ahead of time are still restricted to rehearsing preplanned phrases or situations without interaction from someone who could misunderstand them or play the role of the person with whom they eventually intend to have the conversation.

Sensei_2, who was in the classroom where VocabNomad was deployed, suggested that additional first language scaffolding could be provided and gradually removed. He felt that it may have better supported some of their students’ language-learning needs. While this may be true, student access to translators and their tendency to search for vocabulary within VocabNomad using English rather than Japanese indicates that this may be a teacher perception that needs to be addressed rather than a problem that learners feel exists. The later evaluation of
VocabNomad in informal settings indicates that ELLs do not feel that the application needs to support the translation of vocabulary. They saw value in how it provided support through images, sentences, and definitions. However, those learners were also in an English-language milieu and had become accustomed to non-translation-based learning practices. Learners in both of these studies also had higher levels of vocabulary knowledge and English proficiency than many of those from the initial proof-of-concept study. These differences in user prior knowledge and learning abilities could mean that Sensei_2’s perceptions would apply in a setting where students were less proficient.

While many language-learning activities (e.g., studying and listening or speaking practice) were supported by VocabNomad, learner metacognitive activities were not. From the perspective of users, the system did not appear to track their activities or knowledge because it did not allow them to see this information or perform self-monitoring activities. The need for additional feedback and information was desired by learners, especially those with higher proficiency levels. The desired feedback and support for self-monitoring activities was visible in participant comments about the experience sampling application and the learner reflections on their language-learning processes during the proof-of-concept study. The MALL approach used by VocabNomad provides access to the information that is necessary to support this need through its learner model but this information has yet to be made available to the learner.

9.1.2 RQ2: Integration into Language Learning and Communication Practices

We observed high usage of audio-visual media, translators, and dictionaries across all three of the deployments that were conducted in informal learning environments. This indicates that participating ELLs were comfortable with integrating technology into their language learning and communication practices prior to the introduction of the MALL tool that we provided. This differs from previous results that indicate that students need guidance in how to integrate MALL into their activities (Hockly, 2013).

That said, guidance would still be beneficial since ELLs appear to have a range of experience and facility with using MALL resources. As the participating populations changed, we saw changes in their prior and current use of technologies to support their learning. The participants from the VocabNomad studies in informal learning environments, who were all highly proficient
and college-aged, used more mobile or computer-based applications and fewer tapes or CDs to support their language-learning activities. Their varied tool use and integration of VocabNomad as one of the tools that they employed indicates a need for highly-flexible applications that afford many uses. Many of the participants from the proof-of-concept study used fewer computational resources to support their learning. They also demonstrated less variety and refinement in their learning strategies, which may be related to their sense of self-efficacy or their lack of exposure to technology as a result of their age, socio-economic status, prior education, or profession.

We see a transition in VocabNomad usage that may be linked to participant language proficiency. It appears that higher-level learners were more comfortable integrating VocabNomad into their learning and communication practices as a support tool rather than a primary learning resource or communication device. That is not to say that higher-level learners did not use VocabNomad to study or to support their communication, but that they performed a greater variety of activities using the application than those from the initial proof-of-concept study. The higher-level learners’ use of the system seemed to vary more from participant to participant, much like the already existing learning practices that they reported using. This difference is likely due to these learners having developed the ability to monitor and refine the learning strategies and techniques that they use. This variance in learner strategy, technology, and VocabNomad usage could be partly due to other factors, such as the learner’s home culture or their existing cultures of MALL tool use (Sweeney, 2013). In contrast, the low-level proficiency participants from the proof-of-concept study did not report any self-monitoring activities and only reported using language-learning strategies and technologies that teachers or other learners had told them to use.

To contrast how participants used the applications to support their communication we can consider three examples, each of which comes from a different study. In the proof-of-concept study a low to mid proficiency participant (P) used the text-to-speech feature to help her overcome a communication barrier that resulted from her conversational partner not understanding her accent. In the VocabNomad validation study, a high-proficiency learner from China (FormEv4) used the application to find the specific word that she needed to achieve her goal. To do this, she would have had to search for the appropriate category of items (i.e., vegetables or squash). This would have been followed by visually scanning through the items
until she found the vegetable that she was wanting and obtained its English label. This scanning activity was necessary because the translation of the targeted Chinese word mapped to all pumpkins rather than a particular type of pumpkin. In contrast to these examples of system use for supporting oral language, no one in the final study used the system to support their oral language even though participants had similar proficiency levels and language profiles to those from the VocabNomad validation study. As a result, the third example comes from ESM05 who used the application to support her writing of an essay about her life experiences. The topic of this essay provides evidence for the system’s ability to support interactional communication since it helped her to communicate the type of information that is used to build common ground and enable the development of relationships. These examples show that this type of MALL tool can be used to support both transactional (task-based) and interactional (social) communication. However, social communication may only be supported when ELLs are using written rather than oral communication methods.

The ability to scaffold both forms of communication and participant willingness to use this approach to MALL for scaffolding their communication shows that this tool provides support beyond that typically supplied by communication support applications, which are known to primarily enable transactional communication (Todman & Alm, 2003). However, the difference in usage between the two VocabNomad studies indicates that cultural factors and individual preferences may influence the user’s decision to employ the application as a communication support tool. These individual preferences likely come from the person’s willingness to take risks or make mistakes as well as his or her lack of other strategies for overcoming the communication barrier. Both the MyVoice and squash example provide evidence that participant willingness to use the application to support oral language is at least partly the result of communication failure. This use may indicate that participants lacked the resources or strategies necessary for overcoming the failure and resorted to using the application because they did not know what else to do. Only using the application to support oral communication when a failure has occurred indicates that there are social barriers to using an application for this purpose. These social barriers will be tied to norms and vary from culture to culture, with different social penalties exercised based on the complex negotiation of the identities of those participating in the conversation and those in their immediate surroundings (Brown & Levinson, 1999; Goffman, 1959; Traxler, 2013).
The integration of VocabNomad into a classroom (RQ4) demonstrates how it can be used across educational contexts. Its use across contexts is supported by participant uses of the system during the affect and communication study where they used the MALL tool to complete their homework, much like the students from the Japanese high school did. However, participants in the affect and communication study used VocabNomad when working independently, whereas the high school students used it both independently and in groups even when the activity that they had been given was intended to be completed individually. This shows that the tool can be integrated into both formal and informal environments to support student learning activities.

If we consider a finer-grained view of participant system usage, we see evidence of participants’ desire for socio-collaborative opportunities across two of the three deployment studies that were conducted in informal learning environments. GS created and wanted to contribute vocabulary to the larger community, and participants (FormEv3 and FormEv4) in the VocabNomad validation study wanted to share words with each other. However, the hiding or unsharing of vocabulary items by participants in the affect and communication study indicates that not all participants want to share their learning content; these participant actions justify the decision to allow individual users to control what they share and what they keep private. In addition, user sharing practices indicate that some participants may have been using the system as a personal support tool rather than using it as a generic learning support system, such as Google Translate or dictionary.com.

The detailed logging of participant actions provides support for claims that participants want to be able to organize their learning content (de Jong et al., 2010). While this was not observed in the initial proof-of-concept study, I would argue that the lack of editing activities performed by participants in that study reflects the lack of support that MyVoice provided for on-device editing. The addition of words; editing of vocabulary item content; and addition, removal, or modification of the tags that were associated with vocabulary items during the affect and communication study indicates that participants will exercise their ability to organize learning content when using an application for self-guided study. However, learners in formal education environments may need to be encouraged to take control of their learning and the materials that they are using.
The current community-based curation model appears to be sufficient for use in informal learning environments. However, it falls short in some formal learning environments (RQ4). The teacher-researchers managed to integrate student-generated content into the course using paper-based curation. The use of this process indicates that better support for the curation of content that has been created by students or the larger community should be provided so that teachers can comfortably integrate these types of activities into their classroom. Giving teachers the option of curating student-generated content would better accommodate their integration of the system when their classroom culture does not fully support a socio-cultural approach to learning.

9.1.3 RQ3: Learner Vocabulary Knowledge, Affect, Communicative Success, and System Usage

Participants from the communication and affect study experienced more perceived communicative success when they had access to VocabNomad in spite of the already high levels of communicative success that they were experiencing at the beginning of the study (RQ6 and H4). Their communicative success improved during the study, which may be linked to study instruments and procedures. This success was not always maintained when ELLs no longer had access to VocabNomad, which supports H5. Participating ELLs experienced a reduction in their communication failure when they had access to VocabNomad. By combining this evidence of participant communicative success with ELL reports of using this MALL tool to achieve their communication goals, it can be seen that the use of this adaptive approach to MALL supports participant communication.

Participants in the communication and affect study experienced increased positive affect when they had access to VocabNomad (RQ7), but their positive affect measures appeared to return to baseline levels following system removal, which supports H2 but not H3. Similarly, participants’ minimum level of positive affect improved when they had access to the support tool and returned to baseline levels upon its removal. Improvements in positive affect along with the improved negative affect of some participants indicate that having access to an adaptive MALL support tool improved learner affect in the short term. This may influence later communication and learning practices, such as the observed increase in the number of communication attempts made by participants during the reversal phase. Learning outcomes do not appear to be related to these
affective state improvements since no relationship between VocabNomad usage and ELL vocabulary knowledge was observed during this short-term study.

In contrast, students from the high school study showed gains in their vocabulary knowledge when they were using the developed MALL approach (RQ5 and H1). Consistently high student scores on their final exam further indicate that the tool supports vocabulary development since students learned the words that were distributed through VocabNomad. Students’ self-perceived English-language abilities also improved when they were using the MALL tool; the control group did not show improvements on any of the measures, which supports H1. This difference in language-ability improvement indicates a relationship between longer-term system usage and vocabulary knowledge. The developed decision tree and rules help us to describe the relationships among learner vocabulary knowledge, language abilities, and VocabNomad usage. It should be noted that the resulting models maintain the school’s current status quo for keeping students in the advanced streams once they have been placed there. This is not inherently bad but should be kept in mind when using the models in other settings.

9.2 Contributions

In addition to the smaller contributions that are detailed throughout the discussion of the main research questions, four larger contributions have been made.

The first relies on the use of the detailed tracking upon which the ecological approach is dependent. These detailed logs enabled the development of a model that predicts learning outcomes based on the learner’s language knowledge and MALL tool usage. This model is the first that ties the usage of a MALL tool to changes in learner vocabulary knowledge (RQ3). The methods used to develop the model of MALL usage and its relationship to learning outcomes also serves as an example for other MALL researchers who want to start working on defining and refining the relationship between MALL usage and users’ learning outcomes.

The second contribution centers on the holistic evaluation of a MALL tool from the perspectives of the language input that the tool provides to the learner, changes in learner knowledge, changes in learner communicative success, and changes in the learner’s affective state. All three of these outcome measures have not been studied in combination with the language input that ELLs
receive as the result of MALL usage. Studying the dependent variables of communication, affect, and vocabulary knowledge together has allowed me to detail how different learners use technology to overcome communication barriers and shortcomings in their knowledge. This includes information about how learners have appropriated technologies to scaffold their understanding and fill gaps in their knowledge even though these technologies were not intended to support language-learners. This understanding goes further than previous work by detailing ELL use of mobile tools and highlighting the differences in learner use based on their habits, preferences, and knowledge (RQ2). The MALL tool uses described by learners in my studies also provide examples of concrete tasks that teachers could recommend to students. These examples can be used to help teachers identify the learning strategies and support tools that most effectively meet each student’s needs.

The third contribution is showing that the ecological approach to learner modeling can be used to recommend learning materials in an ill-defined domain, such as the learning and use of language, where the technology-enhanced learning environment infers rather than measures learner knowledge. This contribution includes the demonstration that using the ecological approach to recommend learning content within a mobile adaptive system is accompanied by improvements in learner vocabulary knowledge and communicative success (RQ3). I also show that a system that employs the ecological approach to support learning is accompanied by improvements in the learner’s affective state (RQ3). Systems that have employed the ecological approach have either been evaluated for their ability to support learning using simulated learners and wizard of oz techniques (Champaign & Cohen, 2013; Erickson, Frost, Bateman, & McCalla, 2013) or they have not been evaluated for their effects on learning (Benlamri & Zhang, 2014; Brooks, Winter, Greer, & McCalla, 2004). Other work that relies on the ecological approach has taken initial steps towards its realization or built on it to develop new frameworks (Jovanovic, Gašević, & Devedžić, 2009; Jovanović, Gašević, Tornai, Bateman, & Hatala, 2009; Jovanović, Tornai, Gašević, Bateman, & Hatala, 2008; Lehmann, Hildebrandt, Rensing, & Steinmetz, 2007; Lehmann, Rensing, & Steinmetz, 2008).

The final contribution is the validation of applying information retrieval techniques to Internet-based corpora for generating communication support materials. This reduces the burden of
creating learning content and reaches beyond the educational technology community to that of the assistive and augmentative communication community.

9.3 Future Work and Potential Expansions

The studies that were conducted as part of this project and which are reported in this document identify several opportunities that could be explored. These future directions and potential expansions fall into three categories: educational opportunities which focus more on the teacher and classroom environment, system development opportunities that focus on the learning approaches used within MALL, and research opportunities that present practical and theory building opportunities. The potential for this work to transfer from the realm of research to the everyday activities of language learners is dependent on the continued exploration of the issues that relate to the instructor and students’ abilities to see the potential of these tools and integrate MALL into their everyday practices in a way that benefits ELL needs.

9.3.1 Educational Opportunities

There is a clear need for the development of guidelines for the integration and use of MALL within formal educational environments. There is also a need for the inclusion of the perspectives of language instructors in the development and later uses of MALL tools.

At present, instructors receive limited or no guidance for integrating MALL tools into classroom settings (Graddol, 2014) even when the literature on mobile learning is consulted (Stockwell & Hubbard, 2013). Teachers seem to struggle with the evaluation of MALL tools, which has led to the recent development of checklists for their evaluation (Son, accepted) but not guidelines for their integration.

The work that I conducted confirms the need for activities to progress in difficulty based on the technological abilities and language knowledge of learners (Sweeney, 2013) by advising instructors to adjust this progression based on student familiarity with the specific system and not only the students’ comfort or facility with the devices themselves. This work confirms the need for teachers to have appropriate technical support in their classrooms when attempting to integrate MALL tools (Hockly, 2013). This need may mean that some schools will face greater
barriers to integrating MALL tools as a result of teacher technical knowledge and a lack of support for the technical infrastructure that many MALL tools require.

Beyond making small incremental changes, the most important aspect of the success of VocabNomad’s integration into a high-school classroom was the use of explicit activities to encourage system use. This type of guidance is also necessary for learners with lower levels of metacognitive abilities, especially when they are using MALL in informal learning environments. The development of features or accompanying materials that enabled the tool’s integration into users’ language-learning activities would better support its effective use. These materials and features would also reduce the burden placed on teachers should they want to try integrating the MALL tool into their classrooms.

Our incremental adjustments to the use of VocabNomad in a Japanese high school show how a tool can be integrated into an existing culture of use (Sweeney, 2013) and help to describe that culture of use. This integration helped to challenge the existing culture of use by showing the classroom teachers that students could learn English through English and did not always need to have content translated. While the communicative approach advocates the use of English to learn English and the Japanese ministry have created policies requiring this approach (MEXT (Japanese Ministry of Education), 2012), there has been little adoption. Conversations with a teacher-educator in Japan have explained this lack of uptake as the result of there being a complete absence of teacher training in how to use this approach or technologies for supporting English-language learning (Hajime, 2014).

Some teacher education programs outside of Japan incorporate MALL (Sweeney, 2013), but many do not. Recent investigations into teacher training and the experiences of the teachers from the classroom-based study indicate that programs need to create training that develops teachers’ ability to evaluate, explore, and integrate MALL or other technologies into their classrooms (Beatty, 2013; Son, accepted; Sweeney, 2013). It is not enough to rely on students to perform this task, even if they are highly proficient in the use of technology, because many students lack the other competencies required for integrating technology into the language-learning environment of a classroom (Gajek, accepted). In addition to the development of teacher training, follow up should be performed on how this influences their later integration of MALL
into their classrooms so that best practices for the integration of these technologies and the training of teachers can be developed.

The independent and classroom based learning activities that are supported through VocabNomad could better scaffold the reflective and metacognitive activities of learners. In a classroom setting, this could be done through the use of journal activities. For informal learning environments, this could be accomplished through the use of an open learner model or learning dashboard (Bull & Kay, 2013).

### 9.3.2 MALL Development Opportunities

VocabNomad’s general architecture and use of the ecological approach for tracking learner actions and modeling their knowledge enables the later expansion of system features. Participant interviews and language-learning theory indicate several avenues for improvement in the functionality provided by MALL tools. This includes increased personalization, additional scaffolding, the provisioning of meaningful feedback, and increased support for socio-collaborative activities. The features that are proposed below do not always neatly fall into one of these categories but all of them could be added to the studied tool, provided the appropriate resources were made available.

#### 9.3.2.1 Adaptive Features and System Recommendations

A MALL tool that provides increased personalization holds the potential to improve learning outcomes as the result of its close alignment with dynamic assessment practices and its potential for helping learners to transcend their current abilities by providing the scaffolding that is necessary for them to achieve their potential. Specific, shorter-term goals that could be used to increase the personalization provided via the studied tool are discussed. This list is not exhaustive. Rather, it highlights achievable steps towards a grander vision of highly adaptive MALL systems that can accurately infer learner knowledge and adjust support materials to the learner and his or her current situation.

The performance of the on-demand vocabulary support needs improvement from an algorithmic perspective. The user interface could also be improved to better communicate what the system is doing and manage user expectations. Improvements to this feature could include the predictive
modeling of user content needs based on the previous support requests of other users, the sequence with which similar users accessed vocabulary, or the user’s scores from the adaptive testing feature. These improvements would allow vocabulary collections to be pre-loaded automatically to prevent potential delays in obtaining them based on connectivity and network latency. The preloading of vocabulary could be implemented in a way that allows users to exercise control over whether these collections are added to their account. The exact design of the import process when a future need has been predicted would require study prior to its implementation.

Adding the adaptive testing feature to the mobile application would help enable this process. However, the format of the test would need to be changed so that it is appropriate for use on a smartphone. It would also need to be changed to enable assessment without the intervention of a human. This could be achieved through the use of matching activities, using the vocabulary entry’s definition or picture, or through fill in the blank test items where learners must select or type the correct word to complete a text. The modification of the tests to take into account a vocabulary item’s part of speech or assess different forms of a word based on common morphological variants of previously studied vocabulary could be interesting. Allowing users to define their activities and locations could allow for the later recommendation of materials based on the activities that users typically perform in the locations that they visit. This user input could increase the relevance of learning materials and improve learning.

To better support the user’s extended mapping or bootstrapping processes, adaptive sentence display could be added. This adaptivity would require a slight re-architecting of the system so that sentences could be learning objects on their own rather than a part of a learning object, as they are now. This modification could benefit from recent work on identifying appropriate sentences to support vocabulary learning (Mostow, Gates, Ellison, & Goutam, 2015) and could be used to help increase users’ depth of knowledge for different vocabulary items by showing learners how the word can be used in different contexts.

The above modifications to testing and the extended use of VocabNomad could help improve both the predictive and the descriptive models that were built as part of this thesis. That said, the existing models could be integrated. This integration would require that the self-assessments of
language knowledge and the assessments of morphological knowledge be incorporated within the MALL tool. The self-assessment is short-enough that it could be easily integrated and performed immediately following the elicitation of the learner’s mother tongue while the base learning content is being downloaded. Some questions from the morphological assessment could also be posed at this time. However, those assessments are too long and would need to be divided into smaller chunks that could be asked at regular intervals until all of the items have been completed.

More interestingly, the current model could be integrated as an early-warning system that draws teacher attention to the learners who are not in the mid-performing group. The teacher could then identify those who belong to the low performing and high performing groups and provide additional support to those in the low performing group. Adding a feature that enables this process would help improve the model by adding labels to cases so that it can be retrained and improved.

The information tracked about the editing of vocabulary items could further be used to recommend instructional materials or to modify the existing base and importable vocabulary collections. This modification could be done with a human-in-the-loop approach where items are identified based on user edits and an approved editor reviews and adjusts them. Alternatively, this modification of the base and importable vocabularies could be automated, which would be more difficult than a human-in-the-loop approach.

9.3.2.2 Scaffolding

ELLs want and need additional support. However, we do not want learners to become dependent on a particular tool or other form of support. We want to develop independent learners by adjusting the type of scaffolding that is provided or by appropriately decreasing the amount of support that is given to learners.

Learner preferences for using certain types of media should be supported. However, the reliance on specific types of media to the exclusion of others should be discouraged. This means that learners should be encouraged and perhaps even rewarded for the use of the language modalities (e.g., reading or listening) that they often neglect. This type of individualized scaffolding for the learner’s affective and meta-cognitive needs could be enabled through a simple reminder
mechanism or a more interactive experience where a human or computational agent provides the necessary encouragement and guidance.

The recommendation of instructional materials based on the adaptive testing or the addition of the morphological assessments to the system could be used to support ELLs. The addition of instruction based on the appropriateness of different language usage for a given setting could also help to better support ELLs. This addition would not be technically difficult but would require considerable design and knowledge engineering effort.

The addition of mother tongue scaffolding, in the form of a translation could help learners. Should this support be added, it would be prudent to phase it out based on increases in learner knowledge. The decision to provide translations without the learner having to explicitly request them could be configurable or adaptive when it is performed at a global scale. More interestingly, it could be done at the categorical (i.e., tag) or individual word level based on inferences of the learner’s familiarity with a specific vocabulary item or the vocabulary that are used in its definition.

The user’s ability to plan and rehearse oral communication could be supported through the addition of various features including conversational agents or script writing tools. Script writing tools could include grammar support and suggestions about word collocations. Once the user has recorded a rehearsal attempt, the system could provide feedback based on the user’s pronunciation or intonation.

9.3.2.3 Feedback

Along with scaffolding, the delivery of appropriate feedback is important to learner development. In keeping with socio-cultural theory, the use of open learner models and learning dashboards provides the individualized feedback that learners need to plan and adjust their learning.

Adding an open learner model that provides the user with information about his or her learning activities and knowledge could help learners develop the monitoring strategies that are needed to enable their effective learning. In addition to this, the use of goal setting and gamification could be employed to support self-regulated learning and maintain learner motivation. A brief
exploration into the design and integration of a gamification-based open learner model within VocabNomad showed that that this approach has potential (Tsourounis et al., 2014). However, it requires additional validation and study, especially if we want to help the learner develop his or her meta-cognitive skills.

The recently developed open learner model, would allow the system to better target instructional support materials, assessments, and feedback based on the learner’s goals. Open learner models can also be used to address the need to encourage learners to keep working (Demmans Epp & McCalla, 2011; Stockwell & Hubbard, 2013). Furthermore, one that incorporates gamification and goal setting (Tsourounis & Demmans Epp, in press) would allow us to better understand when gamification works within an open learner model and when it does not since these topics have not been investigated together.

Beyond using an open learner model to communicate learner abilities, the integration of this type of feedback could be used to determine how learners respond to feedback from systems that infer their knowledge rather than measure their knowledge. Because the determination of learner knowledge is inferred in the MALL approach used by VocabNomad, there is more uncertainty in the resulting inferences than there would be in a system that relies on test results to measure learner knowledge. VocabNomad’s reliance on an inference-based approach to determining learner knowledge means that the system is reasoning over information that has increased levels of uncertainty. As a result, the MALL tool may misdiagnose student strengths or weaknesses. The implication from this and Kay’s call for scrutability (2006), which is the ability for the user to understand his or her model, would be for the open learner model to represent the uncertainty that is present as a result of the dynamic, inference-based modelling process. However, we do not know how learners respond to the representation of model uncertainty and the best methods for representing uncertainty within educational reporting is still an open question (Demmans Epp & Bull, 2015; Demmans Epp, Bull, & Johnson, 2014).

9.3.2.4 Socio-Collaborative Support

Participants desired opportunities for performing socio-collaborative learning activities, which are only provided through MALL that employ computer-mediated-communication approaches to supporting user language development. Dedicated MALL tools, where users are not expected to
learn through their interactions with others, fail to address this learner need. As such, the use of socio-collaborative approaches within MALL could be better understood.

The current sharing and importing system could be expanded to better meet learner expectations for this type of learning activity. At a minimum, the ability to browse importable content by user should be added. This would not be a difficult feature to add but was not included because the ability to share learning content between users was not a focus of the current research contribution for this system. The ability to browse importable content by user should likely be accompanied by some form of reputation management where learners can adjust their profile and share information about themselves. Along with individual profiles, the determination and display of one’s relationship to the community of users or position within that community could be added. An explicit curation system, where privileged users are allowed some level of control over community member’s contributions, may better support socio-collaborative activities without threatening content quality. This could be done using a number of processes or mechanisms and user privilege could be based on their contributions to the community or an official status, such as the teacher in formal educational environments.

Enabling knowledge construction through a dedicated MALL tool may enable us to better track and explain how power dynamics influence learning in settings where knowledge and learning materials are co-constructed by different stakeholders across varied educational and cultural contexts.

9.3.3 Research Opportunities

Many possible study opportunities arise out of the developed MALL approach and its proposed expansions. These opportunities include the improvement of individual features and the exploration of this approach’s effect on language learners, their learning environment, and their learning outcomes. These explorations hold the potential to inform language-learning theory.

In addition to the better provisioning of socio-collaborative supports for learning, exploring the use of these features and their impact on communication or language learning would be of interest, especially if conversational agents were incorporated to support the ELL desire for additional communication opportunities.
The longer-term deployment of the system and its integration into a language-learning program would help to validate and further clarify the results that I have reported. The data collected during a study that involved system use over multiple terms would allow the generalizability of the identified models to be tested and may enable the development of better models. The continued use of the tool in different settings may help us to determine why many users treated the application as a content-delivery tool rather than one over which they could exercise agency. Furthermore, the addition of the open learner model that includes goal setting and gamification could be explored for its influence on this phenomenon; system usage; and learner affect, communication, or vocabulary knowledge.

Re-conducting the affect and communication study with a modified protocol, where the baseline is taken during week one of the study and the treatment measures are taken several weeks after the introduction of VocabNomad may help clarify the relationship between system usage and the learner’s affective state, communicative success, and language knowledge.

Ideally each of these deployment studies, whether they focus on vocabulary, communication, or affect would be performed with larger sample sizes and be repeated with different populations to better understand the variables involved. The use of detailed tracking and experience sampling within these evaluations may enable us to identify the relationships between two models of foreign language education: the input-interaction-output model and the socioeducational model. These language-learning models are treated independently (Dixon et al., 2012) even though there is evidence that the variables from one model influence the constructs studied within the other model. Continuing to study the use of a MALL tool from the multiple perspectives that were employed throughout this thesis will allow the required data to be collected so that these models can be linked and possibly even unified.

The proposed replication would also help us to better understand when this adaptive MALL approach is useful for supporting each of these factors based on the user’s language profile. An increased understanding of when this approach is useful would allow us to determine the thresholds that limit VocabNomad’s usefulness for supporting vocabulary learning or communication. In addition to replication, the extended use of the system could enable the determination of how learner familiarity affects system usage or how continued use influences
user’s language learning and communication strategies. Of particular interest, would be the
determination of how learner strategies for handling communication failures change with the
introduction and continued use of the technology.

The detailed exploration of tag edits could lead to a better understanding of how the schemata of
ELLs from different backgrounds develop. Exploring tag editing may help inform our
understanding of how learners’ depth of knowledge of different vocabulary develops. It could
even lead to our ability to create appropriate scaffolds and instructional materials to support their
schemata and vocabulary knowledge development by identifying where ELLs struggle with
respect to the multiple meanings of words.

Finally, a study exploring how teacher attitudes and actions influence system usage by students
would provide interesting insight into how these types of applications can be integrated into
formal learning environments. This is not yet well understood (Woodman, accepted) and could
help system designers to overcome some of the barriers that prevent the adoption of these tools
in formal educational environments.

9.4 Summary

The different ways in which ELLs used VocabNomad show that it is consistent with many of the
recently-developed principles for MALL (Stockwell & Hubbard, 2013; West & Vosloo, 2013)
and the design recommendations that were developed in Chapter 3. The approach used by
VocabNomad pushes content to learners but respects their autonomy through its content
recommendation and dismissal feature. It accommodates ELL differences by providing
personalized recommendations and allowing the user to exercise control over his or her learning
through its translation feature and its vocabulary search and import feature. VocabNomad was
also designed in a way that allows it to be adapted to meet varied cultures of practice. The
realization of this design accommodation is demonstrated through its use as a support tool within
class and its use for communication support, vocabulary review, self-testing, or listening and
pronunciation practice outside of class. The learning tasks, as represented by the largest of the
learning objects, are short and small. Learners control their appropriation and use of
VocabNomad’s learning activities based on the spaces that they occupy. The system also
incorporates basic training in the form of instructions that are visual or textual. We refined prior
recommendations for training learners to specify incremental and repeated training because basic training at the beginning of a course does not appear to be sufficient for ensuring full system use and integration when working in formal educational environments.

In addition to meeting many of the UNESCO guidelines and other recently developed principles for MALL, VocabNomad is one of only a few systems that employs the ecological approach to learner modeling and provides evidence for its effective use in the wild (Demmans Epp, 2015; Demmans Epp et al., 2013). This form of learner modeling and the integrated content generation and recommendation distinguishes VocabNomad’s approach from that of other MALL tools as described by Burston (2014a) and Traxler (2013). The level of detailed tracking of learner actions that is required to enable the ecological approach also distinguishes this approach to MALL which is evidenced through calls for logging (Ma, 2013) and a survey of MALL tools and use (Burston, 2014a). Participant use of the approach to MALL that I proposed and operationalized within VocabNomad demonstrates the ability of this approach to support learner-driven activities in informal environments and further differentiates this work. This MALL tool is also one of the few that was designed to support language learning by enabling noticing through its on-device editing and content creation feature as well as its audio recording feature. Beyond these specific technological approaches and features, its demonstrated ability to support learner-driven activities alongside its ability to support teacher-directed activities that are not completely specified is not commonly observed (Kukulska-Hulme, 2013), nor is the successful integration of one system into learner practices in both formal and informal learning environments.

Thus far, there has been little evidence that MALL tools provide communication or pastoral support (Traxler, 2013). My work shows that MALL can be used to support both oral and written communication in real world settings. It also shows that the proposed approach to MALL provides pastoral support by scaffolding ELL communication as exemplified by FormEv4’s use of VocabNomad to find a particular type of squash at the grocery store. My work confirms prior arguments for the ability of MALL to support transactional communication and provides an example of MALL tool use that supports interactional communication in written form.
The data collected during the affect and communication study (Chapter 8) extend our current understanding of ELL strategies when using technology as elaborated by Kukulska-Hulme (2013) and Viberg and Grönlund (2013), by detailing when and how ELLs employ certain technologies to support their comprehension. My data does not support Beatty’s (2013) claim that mobile tools are supplanting more traditional approaches to language learning. The reality appears to be far more complex with learners integrating mobile tools, computer-based programs, courses, text-based media, and other technologies to support their learning. The detailed logging of user actions that is required of the ecological approach enabled the further description of learner strategies around the use of an adaptive MALL tool that aims to support the communication and language learning activities of ELLs. Participant reports and observational data from multiple studies show learners performing more collaboration than is often reported (Sweeney, 2013): participants were observed working in groups, editing the materials that they had studied, and discussing the vocabulary that they had created or edited and then shared.

The vocabulary study of VocabNomad was considerably longer than the majority of those that have been reported (Burston, 2014b). Student usage patterns provide evidence of the novelty effect, which could not have been observed in the majority of prior studies; prior studies did not last long enough for the systems' novelty to have worn off.

The repeated positive results from studies that were conducted across learning and cultural contexts, with different types of learners, distinguishes this work from that of others which often lacks reliability or has only been conducted in a single educational and cultural context (Hockly, 2013; Stockwell & Hubbard, 2013). Both summative studies included measures of language knowledge, some of which were already established. This application of language assessment and usage tracking is rare in studies of MALL (Burston, 2014a; Traxler, 2013). The collection of these measures enabled the creation of models of system usage and its relationship to changes in learner knowledge. The development of these types of models has not been previously reported, even though some differences in how learners approach using MALL have been reported (Stockwell & Hubbard, 2013).
The evaluations conducted as part of this research project have been performed with populations whose MALL use is not typically explored. This includes its use with marginalized populations (e.g., recent immigrants), second rather than foreign language learners, high school students, and advanced language-learners (Burston, 2014b; Kukulska-Hulme, 2013; Traxler, 2013). As a result, the findings from these studies expand our knowledge of ELL use of MALL for underserved populations.

The contributions that are made within this thesis centre on the MALL approach that is operationalized through VocabNomad, its evaluation contexts, and the evaluation of learner affect, communication, and knowledge. The MALL approach that was proposed and implemented is the first to employ the ecological approach to learner modeling. The detailed tracking that the ecological approach requires enabled the development of models of student application use and vocabulary knowledge that were not previously possible because prior studies did not include detailed tracking of learner activities within MALL tools. Learner reports of their use of VocabNomad and others systems expands our knowledge of how mobile tools are used to support varied language-learning activities based on user needs. These needs include ELL communication, affect, and vocabulary knowledge. VocabNomad’s use by language learners who have different language profiles and backgrounds also broadens our knowledge of how ELLs can benefit from the use of these types of tools.

The above described contributions are spread throughout the formative investigations (Chapter 3 and Chapter 4) that resulted in the implementation of an adaptive approach to MALL (Chapter 5) that was evaluated for its influence on learner knowledge, communication, and affect (Chapter 7 and Chapter 8). The first of the formative evaluations showed that this type of system could be used to support ELL communication and language learning activities. This evaluation also resulted in a set of design recommendations that were used to inform the development of a MALL tool. The tool’s user interface underwent several forms of validation and the content-generation that was integrated into this tool was validated before VocabNomad was tested in a final formative evaluation. VocabNomad was then evaluated in two summative assessments. These studies showed that VocabNomad’s approach to MALL can support vocabulary learning. They also showed that users experience greater communicative success and improved affect when they have access to the MALL tool.
The work conducted here shows the benefits of using systems that employ the ecological approach to support user communication and learning. This work opens several avenues for future exploration. One of which is the necessary replication of these studies to confirm their results and improve the accuracy of the developed models. Studying the continued use of the system by users would also help us to better understand how their practices change based on system familiarity and their changing language profiles. The identified development opportunities present design and technical challenges that could help to better support learning. That said, the approach to MALL that is proposed, implemented, and evaluated throughout this thesis currently meets the needs of a variety of language learners. It can support their communication, their affective state, and the development of their vocabulary knowledge.
References


Ballance, O. J. (2013). MALL—Somewhere between the tower, the field, the classroom, and the market: A reply to professor Stockwell’s response. Language Learning & Technology (LLT), 17(1), 37–46.


Beatty, K. (2010). *Teaching and researching computer-assisted language learning* (2. ed.). Harlow [u.a.]: Longman.


226


Hawkins, W., Heffernan, N., & Baker, R. S. J. D. (2013). Which is more responsible for boredom in intelligent tutoring systems: Students (trait) or problems (state)? (pp. 618–623). IEEE. http://doi.org/10.1109/ACII.2013.108


Kukulska-Hulme, A. (2013). *Re-skilling language learners for a mobile world* (pp. 1–16). Monterey, CA, USA: The International Research Foundation for English Language


Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, 20(03), 271–289. http://doi.org/10.1017/S095834408000335


http://doi.org/10.1080/09588221.2014.987035


http://doi.org/10.1111/j.1365-2729.2009.00329.x


Surfaces (pp. 437–440). New York, NY, USA: ACM.
http://doi.org/10.1145/2512349.2514922


242


Woodman, K. (accepted). Strategies for supporting teachers in the implementation of MALL. In Palalas, Agnieszka & Ally, Mohamed (Eds.), *The International Handbook of Mobile-Assisted Language Learning*.

Appendix A – Ethics Approval

A.1 Formative Evaluation (#26592)

PROTOCOL REFERENCE # 26592
July 22, 2011
Dr. Ronald M. Baecker
Department of Computer Science
University of Toronto
40 St. George St.
Toronto, ON M5S 2E4

Ms. Carrie Demmans Epp
Department of Computer Science
University of Toronto
40 St. George St.
Toronto, ON M5S 2E4

Dear Dr. Baecker and Ms. Demmans Epp:

ETHICS APPROVAL
Original Approval Date: July 22, 2011
Expiry Date: July 21, 2012
Continuing Review Level: 1

We are writing to advise you that the Social Sciences, Humanities and Education Research Ethics Board has granted approval to the above-named research study under the REB’s delegated review process. Your study has been approved for a period of one year and ongoing projects must be renewed prior to the expiry date.

All your most recently submitted documents have been approved for use in this study.

Any changes to the approved protocol or consent materials must be reviewed and approved through the amendment process prior to its implementation. Any adverse or unanticipated events should be reported to the Office of Research Ethics as soon as possible.

Please ensure that you submit an Annual Renewal Form or a Study Completion Report 15 to 30 days prior to the expiry date of your study. Note that annual renewals for studies cannot be accepted more than 30 days prior to the date of expiry, as per federal and international policies.

If your research has funding attached, please contact the relevant Research Funding Officer in Research Services to ensure that your funds are released.

Best wishes for the successful completion of your project.

Yours sincerely,

Dean Sharpe, Ph.D.
Research Ethics Board Manager–Social Sciences and Humanities

OFFICE OF RESEARCH ETHICS
McMurtrie Building, 12 Queen’s Park Crescent/West, 2nd Floor, Toronto, ON M5S 1S8 Canada
Tel: +1 416 946-3273 • Fax: +1 416 946-3563 • ethics.review@utoronto.ca • http://www.research.utoronto.ca/for-researchers/administrators/ethics/
PROTOCOL REFERENCE # 26431
May 31, 2011

Dr. Ronald M. Baecker
Department of Computer Science
University of Toronto
40 St. George St.
Toronto, ON M5S 2E4

Ms. Camie Demmams Epp
Department of Computer Science
University of Toronto
40 St. George St.
Toronto, ON M5S 2E4

Dear Dr. Baecker and Ms. Epp:

Re: Your research protocol entitled, "Vocabulary Creation and Analysis for Context-Sensitive Vocabulary Support"

ETHICS APPROVAL
Original Approval Date: May 31, 2011
Expiry Date: May 30, 2012
Continuing Review Level: 1

We are writing to advise you that the Social Sciences, Humanities and Education Research Ethics Board has granted approval to the above-named research study under the REB’s delegated review process. Your study has been approved for a period of one year and ongoing projects must be renewed prior to the expiry date.

All your most recently submitted documents have been approved for use in this study.

Any changes to the approved protocol or consent materials must be reviewed and approved through the amendment process prior to its implementation. Any adverse or unanticipated events should be reported to the Office of Research Ethics as soon as possible.

Please ensure that you submit an Annual Renewal Form or a Study Completion Report 15 to 30 days prior to the expiry date of your study. Note that annual renewals for studies cannot be accepted more than 30 days prior to the date of expiry, as per federal and international policies.

If your research has funding attached, please contact the relevant Research Funding Officer in Research Services to ensure that your funds are released.

Best wishes for the successful completion of your project.

Yours sincerely,

[Signature]

Dean Sharpe, Ph.D.
Research Ethics Board Manager—Social Sciences and Humanities

OFFICE OF RESEARCH ETHICS
McMaster Building, 111 Queen’s Park Crescent West, 2nd Floor, Toronto, ON M5S 1A8 Canada.
Tel: +1 416 946-3273 • Fax: +1 416 946-3363 • ethics.review@utoronto.ca • http://www.research.utoronto.ca for researchers/administrators/ethics/
PROTOCOL REFERENCE # 26515
May 28, 2011

Dr. Ronald M. Baecker
Department of Computer Science
University of Toronto
40 St. George St.
Toronto, ON M5S 2E4

Ms. Carrie Demmans Epp
Department of Computer Science
University of Toronto
40 St. George St.
Toronto, ON M5S 2E4

Dear Dr. Baecker and Ms. Epp:

Re: Your research protocol entitled, "Image Identification, Association, and Analysis for Context-Sensitive Vocabulary Support"

ETHICS APPROVAL

Original Approval Date: May 28, 2011
Expiry Date: May 27, 2012
Continuing Review Level: 1

We are writing to advise you that the Social Sciences, Humanities and Education Research Ethics Board has granted approval to the above-named research study under the REB’s delegated review process. Your study has been approved for a period of one year and ongoing projects must be renewed prior to the expiry date.

All your most recently submitted documents have been approved for use in this study.

Any changes to the approved protocol or consent materials must be reviewed and approved through the amendment process prior to its implementation. Any adverse or unanticipated events should be reported to the Office of Research Ethics as soon as possible.

Please ensure that you submit an Annual Renewal Form or a Study Completion Report 15 to 30 days prior to the expiry date of your study. Note that annual renewals for studies cannot be accepted more than 30 days prior to the date of expiry, as per federal and international policies.

If your research has funding attached, please contact the relevant Research Funding Officer in Research Services to ensure that your funds are released.

Best wishes for the successful completion of your project.

Yours sincerely,

Dean Sharpe, Ph.D.
Research Ethics Board Manager—Social Sciences and Humanities

OFFICE OF RESEARCH ETHICS
McMurrich Building, 12 Queen's Park Crescent West, 2nd Floor, Toronto, ON, M5S 1A5 Canada
Tel: +1 416 946-3273 • Fax: +1 416 946-8703 • ethics.research@utoronto.ca • http://www.research.utoronto.ca/for-researchers-administrators/ethics/
PROTOCOL REFERENCE # 29881

February 13, 2014

Dr. Ronald Baeccker  Ms. Carrie Demmans Epp
DEPT OF COMPUTER SCIENCE DEPT OF COMPUTER SCIENCE
FAC OF ARTS & SCIENCE FAC OF ARTS & SCIENCE

Dear Dr. Baeccker and Ms. Carrie Demmans Epp,

Re: Your research protocol entitled, "Using VocabNomad in high school English classes"

ETHICS APPROVAL

Original Approval Date: February 13, 2014
Expiry Date: February 12, 2015
Continuing Review Level: 1

We are writing to advise you that the Social Sciences, Humanities, and Education Research Ethics Board (REB) has granted approval to the above-named research protocol under the REB’s delegated review process. Your protocol has been approved for a period of one year and ongoing research under this protocol must be renewed prior to the expiry date.

Any changes to the approved protocol or consent materials must be reviewed and approved through the amendment process prior to its implementation. Any adverse or unanticipated events in the research should be reported to the Office of Research Ethics as soon as possible.

Please ensure that you submit an Annual Renewal Form or a Study Completion Report 15 to 30 days prior to the expiry date of your current ethics approval. Note that annual renewals for studies cannot be accepted more than 30 days prior to the date of expiry.

If your research is funded by a third party, please contact the assigned Research Funding Officer in Research Services to ensure that your funds are released.

Best wishes for the successful completion of your research.

Yours sincerely,

Sarah Wakefield, Ph.D.  Dean Sharpe
REB Chair  REB Manager
A.5 Affect and Communication Study (#28964)

PROTOCOL REFERENCE # 28964
June 4, 2013

Dr. Ronald Baecker
DEPT OF COMPUTER SCIENCE
FAC OF ARTS & SCIENCE

Ms. Carrie Demmans Epp
DEPT OF COMPUTER SCIENCE
FAC OF ARTS & SCIENCE

Dear Dr. Baecker and Ms. Carrie Demmans Epp,

Re: Your research protocol entitled, "Assessing the effect of mobile vocabulary support on the communication and affect of English language learners"

ETHICS APPROVAL

Original Approval Date: June 4, 2013
Expiry Date: June 3, 2014
Continuing Review Level: 1

We are writing to advise you that the Social Sciences, Humanities, and Education Research Ethics B has granted approval to the above-named research protocol under the REB's delegated review process. Your protocol has been approved for a period of one year and ongoing research under this protocol must be renewed prior to the expiry date.

Any changes to the approved protocol or consent materials must be reviewed and approved through the amendment process prior to its implementation. Any adverse or unanticipated events in the research should be reported to the Office of Research Ethics as soon as possible.

Please ensure that you submit an Annual Renewal Form or a Study Completion Report 15 to 30 days prior to the expiry date of your current ethics approval. Note that annual renewals for studies cannot be accepted more than 30 days prior to the date of expiry.

If your research is funded by a third party, please contact the assigned Research Funding Officer in Research Services to ensure that your funds are released.

Best wishes for the successful completion of your research.

Yours sincerely,

Sarah Wakefield, Ph.D.
REB Chair

Dean Sharpe
REB Manager

250
Appendix B – Demographics Forms

B.1 MyVoice Study

1. What is your birth date? ____________ (Day Month Year)
2. What is your mother tongue (the first language that you learned)?

3. What language(s) do you speak at home? ____________
4. What other languages do you speak and how well do you speak them?

<table>
<thead>
<tr>
<th>Language</th>
<th>Proficiency (poor, good, excellent, fluent, ...)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B.2 Formative Evaluation

1. What is your birth date? ____________ (Day Month Year)
2. When did you arrive in Canada? ____________ (Day Month Year)
3. Are you Male or Female?
4. What is your mother tongue (the first language that you learned)?

5. What language(s) do you speak at home? ____________
6. Which other English language environments have you lived in?

<table>
<thead>
<tr>
<th>Place</th>
<th>Start Date (Day Month Year)</th>
<th>End Date (Day Month Year)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. What other languages do you speak and how well do you speak them?

<table>
<thead>
<tr>
<th>Language</th>
<th>Please circle one per skill 0 = None 10 = Perfect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speaking</td>
</tr>
<tr>
<td></td>
<td>Understanding</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th>Please circle one per skill 0 = None 10 = Perfect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speaking</td>
</tr>
<tr>
<td></td>
<td>Understanding</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
</tr>
</tbody>
</table>
B.3 Vocabulary Study & Communication and Affect Study

The only difference between the forms used during the classroom-based vocabulary study and the communication study was the addition of the questions about participant sex and their arrival in Canada (number 2 and 3) for the communication and affect study.

1. What is your birth date? _______________ (Day Month Year)
2. When did you arrive in Canada? _______________ (Day Month Year)
3. Are you Male or Female?
4. What is your mother tongue (the first language that you learned)?
5. What language(s) do you speak at home? _______________
6. Do your parent(s) know English? Yes / No
   a. Which parent?
   b. How well?
   c. How much do you use English with him/her?
7. Does anyone else in your house speak English?
   a. Who?
   b. How well?
   c. How much do you use English with him/her?
8. Which English language environments have you lived in or visited?

<table>
<thead>
<tr>
<th>Place</th>
<th>Start Date (Day Month Year)</th>
<th>End Date (Day Month Year)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Which of the following English materials do you use outside of school?

<table>
<thead>
<tr>
<th>Material</th>
<th>Amount Used (e.g., times/week)</th>
<th>Other Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td>Sub-titled</td>
<td></td>
</tr>
<tr>
<td>Movies</td>
<td>Sub-titled</td>
<td></td>
</tr>
<tr>
<td>Books</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magazines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Music</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Games</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. What English-language learning activities do you perform outside of school?
11. What English-language learning activities have you performed outside of school?
12. What other languages do you speak and how well do you speak them?

<table>
<thead>
<tr>
<th>Language</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proficiency Level</strong></td>
<td><strong>Please circle one per skill 0 = None, 10 = Perfect</strong></td>
</tr>
<tr>
<td>Speaking</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
<tr>
<td>Understanding</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
<tr>
<td>Reading</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th>Mother Tongue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proficiency Level</strong></td>
<td><strong>Please circle one per skill 0 = None, 10 = Perfect</strong></td>
</tr>
<tr>
<td>Speaking</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
<tr>
<td>Understanding</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
<tr>
<td>Reading</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proficiency Level</strong></td>
<td><strong>Please circle one per skill 0 = None, 10 = Perfect</strong></td>
</tr>
<tr>
<td>Speaking</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
<tr>
<td>Understanding</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
<tr>
<td>Reading</td>
<td>None 0 1 2 3 4 5 6 7 8 9 10 Perfect</td>
</tr>
</tbody>
</table>

253
Appendix C– Interview Guides

C.1 MyVoice Study

1. What has your experience with learning languages been like?
   - Which languages have you tried to learn?
   - Why did you decide to learn that language?
   - What is a typical day like for you in that language?
   - How did you go about learning the language?
   - What types of things help you with learning languages? Why?
   - What helped you the most with learning the language? Why?
   - What tools and strategies did you use? Why?
   - How did they help you?
   - How did they frustrate you?
   - Was there anything that you felt was missing that might have been helpful to you?
   - How much do you want to rely on your mother tongue when learning a language? Why?

2. What has your experience with technology been like?
   - What technologies have you used? (computer, mobile phone, VCR, TV, robot, ...)
   - Where did you use that technology? (home, the library, work, ...)
   - How did you use that technology?
   - What does that technology let you do that you couldn't do before?
   - What does that technology prevent/stop you from doing?
   - What does it make easier?
   - What does it make harder?
   - Why do you keep using that technology?
   - Do you have an example of when you liked using it? What happened?
   - Why did you stop using that technology?
   - Do you have an example of when you hated using it? What happened?

3. Before using MyVoice, what was your experience with using computer programs and mobile devices (e.g., iPhone, Android, iPad) for language learning like?
   - How did you learn to use them?
   - How have they helped you?
   - How have they frustrated you?
   - How have you used them to support your learning?
   - Do you have an example of something that you have changed about how you study/learn since using these programs/devices? What is it?
   - Why do you still use some of them?
   - Why did you stop using others?

4. What would you like to see added to these programs and technologies to make learning easier for you?
   - What would you say to another language learner about these programs and technologies?
   - If there was one thing that could be added to make you recommend these programs and technologies to others, what would it be?
• Describe a situation where you really needed language support but didn't have it. What do you think would have helped you when this happened?

5. What has your experience with using MyVoice for language learning been like?
• What have you done with it?
• How has it helped you?
• How has it frustrated you?
• How have you used it to support your learning?

C.2 Formative Evaluation/VocabNomad Validation

The topics that were covered in the MyVoice protocol were also discussed during the VocabNomad validation study with the exception that the technology-specific questions were about VocabNomad rather than MyVoice. Additional topics were also discussed.

1. What has your experience speaking with other in English been like?
• What is a typical day like for you in English?
• What makes it easier?
• What makes it harder?
• Do you have an example of a conversation that went really well? What happened?
  What did you do? What did the other person(s) do?
• Do you have an example of a conversation that went poorly? What happened?
  What did you do? What did the other person(s) do?
• Do you have an example of a time when you were misunderstood? What happened?
  What did you do? What did the other person(s) do?
• Do you have an example of a time when you misunderstood someone? What happened?
  What did you do? What did the other person(s) do?

4. How did you use VocabNomad when you were communicating with others?
• Can you think of an example where you used it when talking with someone in English?
  What happened? What did you do? How did the other person respond?
• Did you ever show VocabNomad to someone to help them understand you?

5. How is talking with people in English different when you have VocabNomad compared to now/before?
• How much did you use VocabNomad when you were talking with people?
• How much did you try to use VocabNomad when you were talking with people?
• What worked? What didn’t?

8. How is learning in English different when you have VocabNomad compared to before/now?
• How much did you use VocabNomad to try and learn?
• How much did you try to use VocabNomad to try to learn?
• What worked? What didn’t?

9. Describe a situation where you asked for feedback or have been given feedback about your English language skills (i.e., vocabulary, communication skills).
• Who did you get this feedback from? (e.g., teachers, tutors, friends, family, peers, etc)
• What did you find helpful?
• What didn’t you find to be helpful?
• Is there anything you wanted, but didn’t get? What was it? Why do you think that is?

10. What are some ways that you keep track of your progress when learning a language? (i.e., reflection, planning, self-monitoring, tests).
   • How do you organize yourself when learning?
   • What is your planning process like?

11. What are your experiences collaborating and learning from others? (i.e., compare work with each other, motivate yourself to do as well as high performers)
   • What kind of information do you share?
   • What do you find use when comparing your work with others?
   • What don’t you find useful when comparing your work with others?
   • What are some experiences you may have competing against others?

C.3 Communication and Affect Study

C.3.1 Pre-VocabNomad

1. What has your experience speaking with others in English been like?
   • What is a typical day like for you in English?
   • What makes it easier?
   • What makes it harder?
   • Do you have an example of a conversation that went really well? What happened? What did you do? What did the other person(s) do?
   • Do you have an example of a conversation that went poorly? What happened? What did you do? What did the other person(s) do?
   • Do you have an example of a time when you were misunderstood? What happened? What did you do? What did the other person(s) do?
   • Do you have an example of a time when you misunderstood someone? What happened? What did you do? What did the other person(s) do?

2. What has your experience with learning English been like?
   • Why did you decide to learn English?
   • How did you go about learning English?
   • What types of things help you with learning English? Why?
   • What helped you the most with learning English? Why?
   • What tools and strategies did you use? Why?
   • How did they help you?
   • How did they frustrate you?
   • Was there anything that you felt was missing that might have been helpful to you?
   • How much do you want to rely on your mother tongue when learning a language? Why?

C.3.2 Post-VocabNomad

1. How did you use VocabNomad when you were communicating with others?
1. Can you think of an example where you used it while talking with someone in English? What happened? What did you do? How did the other person respond?
2. Did you ever show VocabNomad to someone to help them understand you?

2. How is talking with people in English different when you have VocabNomad compared to now/before?
   - How much did you use VocabNomad when you were talking with people?
   - How much did you try to use VocabNomad when you were talking with people?
   - What worked? What didn’t?
   - How much did you try to use VocabNomad after we took it away?

3. What has your experience with using VocabNomad for language learning been like?
   - What have you done with it?
   - How has it helped you?
   - How has it frustrated you?
   - How have you used it to support your learning?
   - Do you have an example of when you used VocabNomad to help you learn English? What did you do?
   - Do you have an example of something that you have changed about how you study/learn since using VocabNomad? What is it?

4. How is learning in English different when you have VocabNomad compared to before/now?
   - How much did you use VocabNomad to try and learn?
   - How much did you try to use VocabNomad to try and learn?
   - What worked? What didn’t?
   - How much did you try to use VocabNomad after we took it away?

C.3.3 Post-Reversal

1. What has your experience speaking with others in English been like?
   - What has become easier?
   - What has become harder?
   - Do you have an example of a conversation that went really well? What happened? What did you do? What did the other person(s) do?
   - Do you have an example of a conversation that went poorly? What happened? What did you do? What did the other person(s) do?
   - Do you have an example of a time when you were misunderstood? What happened? What did you do? What did the other person(s) do?
   - Do you have an example of a time when you misunderstood someone? What happened? What did you do? What did the other person(s) do?

2. How did you use VocabNomad when you were communicating with others?
   - Can you think of an example where you used it while talking with someone in English? What happened? What did you do? How did the other person respond?
   - Did you ever show VocabNomad to someone to help them understand you?

3. How is talking with people in English different when you have VocabNomad compared to now/before?
   - How much did you use VocabNomad when you were talking with people?
   - How much did you try to use VocabNomad when you were talking with people?
• What worked? What didn’t?
• How much did you try to use VocabNomad after we took it away?

4. How is learning in English different when you have VocabNomad compared to before/now?
• How much did you use VocabNomad to try and learn?
• How much did you try to use VocabNomad to try and learn?
• What worked? What didn’t?
• How much did you try to use VocabNomad after we took it away?
Appendix D – Image Attributions & Permissions

VocabNomad content uses images from the below sources.

- The Noun Project
- The Tango Icon Library
- The Crystal Project
- Crystal Clear
- PublicDomainPictures.Net
- Webshots
- WPClipart
- CAPL: Culturally Authentic Pictorial Lexicon

In addition to this, the image of the layered approach was taken from the “Layered Evaluation of Interactive Adaptive Systems: Framework and Formative Methods” (Paramythis et al., 2010). Permission to use this image was obtained from Judith Masthoff.
Appendix E – Algorithms

All the algorithms require two steps: the first is the retrieval of information from the Internet and the second is the processing of the information to generate lists of words or phrases. This section outlines the algorithms used for the generation of word lists that are related to an identified topic.

Information is retrieved from the Internet-based corpora by first collecting or scraping the content from the targeted corpus. The scraper was implemented in Java and starts by opening a URL and creating a copy of the website as a raw HTML document. The content is then cleaned: the text-based content that would be displayed to a viewer of the webpage is retained and other meta-data is discarded. Unless otherwise noted, HTML Tidy (Raggett, 2004) is used to clean the retrieved HTML. The retained text is then processed using one of the below algorithms.

E.1 Wikipedia

Corpus: the University of South Florida’s (USF) Free Association Database (Nelson, McEvoy, & Schreiber, 1998) and a cleaned version of a Wikipedia page for an identified topic (“Wikipedia,” n.d.). This page will be called wiki. The USF database contains information about the associations that people make between different words. Words within the USF database are assigned a score based on how frequently people associated the two words.

Figure E.1 shows the pseudo-code for the algorithm: HTMLCleaner was used to clean the Wikipedia content (Wilson, Moore, & vnikic, 2006)

E.2 Dictionary

Corpora: thefreedictionary.com, idioms.thefreedictionary.com, thesaurus.com, and about.com. Only the top search result from about.com was used. About.com will be referred to as the aboutCorpus, thesaurus.com and thefreedictionary.com will be referred to as dictionaryCorpus, and idioms.thefreedictionary.com will be called the idiomCorpus.

Mechanize (Aas, Hylton, Lester, & Lee, 2010) was used to perform dictionary searches through the web forms that are part of the dictionary-based sites.
content <- wiki

usf <- [,] #word, USF score
usf <- IdentifyNouns(content.firstSentence)
usf.append(USF_FreeAssociationDatabase.lookup(usf))
usf.normalizeScore()

phrases <- [,] #word/phrase, Frequency Count
phrases <- getLinkNames(content)
phrases.append(subdivideIntoNgrams(content))
phrases.remove(stopWords)
phrases.FreqCount <- phrases.countFrequency
phrases.NormalizeScore()

results <- usf.append(phrases)
results.sortDescByScore()

**Figure E.1 The Wikipedia pseudo-code.**

The algorithm (Figure E.2) relies on the following algorithms: term frequency – inverse document frequency (tf-idf) (Salton & McGill, 1986), relative location of words (Carpena et al., 2009), and collocation identification (“WordHoard - Finding Multiword Units,” n.d.).

### E.3 Review Sites

Corpus: comments from review sites that include blogto, yelp, googleplaces, urbanspoon, restaurantica, chickadvisor, ourfaves, or tripadvisor.

Term-frequency (tf) was the general algorithm that was used (Salton & McGill, 1986). This was augmented with WordNet (Princeton University, 2010). The function

content <- dictionaryCorpus

results <- [,] #word, score
results <- tf_idf(dictionaryCorpus)
results.sortDesc()
med <- calculateMedianScore(results)
results <- retainResultsScoreGreaterThanOrEqualTo(med)

content <- aboutCorpus
results.append(relativeLocation(content))

phrases <- idiomCorpus
phrases.append(collocation(content))
phrases.remove(containsDuplicateWord)
phrases.remove(stopWords)
phrases.sortDesc()
results.append(tf_idf(phrases))

**Figure E.2 The Dictionary pseudo-code.**
content <- corpus
threshold <- floor(log(content.length())) - 2
content.remove(stopWords)
results <- tf(content)
results.combineWordsUsingLemma()

phrases <- identifyNgrams(content, 1, 5)  # input, min n, max n
phraseResults <- tf(phrases)
results.append(phraseResults)
results <- retainResultsGreaterThanEqual(threshold)

Figure E.3 3The Review Sites pseudo-code.

combineWordsUsingLemma() would take sit, sat, and sitting and combine them into a single entry since they all share a common lemma of sit. The frequency scores of these words would be summed and a single vocabulary entry would result. Figure E.3 shows the pseudo-code for this content generator.

E.4 Website

Corpus: the website that is associated with a location or theme. For example, www.starbucks.com would be the corpus if the user were in a Starbucks coffee shop.

Starting at the homepage, all of the text-based documents that are linked from the page are retrieved. This process is repeated with all of the pages that are linked until a specified link depth is reached. The limit on link depth helps ensure that the corpus is of a tractable size for processing. The text of each of the retrieved documents forms the corpus. If the URL is a PDF file, it is processed using PDFBox (Apache PDFBox, n.d.), which interprets the PDF and returns its content as a String. Additional details are provided in Figure E.4.

maxDepth = 3
content <- retrievePages(corpus, maxDepth)

terms <- content.findNouns()
terms <- identifyNgrams(content, 1)  # input, min size

results <- tf(content, terms)  # input, terms to be counted
results.remove(stopWords)
results.sortDesc()

Figure E.4 The Website pseudo-code.
Appendix F – Discourse Completion Tasks

Table F.1 shows the tasks that were used and the type of communication that they were intended to elicit. Transactional communication (Trans. in Table E.1) aims to achieve a particular goal, while interactional communication (Inter. In Table E.1) is used to support relationships. However, the boundary separating these two forms of communication is sometimes blurred. The type of speech task is also specified, where requests involve asking for something, negotiation involves reaching some level of agreement with the interlocutor or resolving a misunderstanding, and responses involve providing information. The location or general theme under which each of the discourse tasks were completed is also specified.

Table F.1 The discourse tasks, the speech and communication types, and the context for which they were designed to be used.

<table>
<thead>
<tr>
<th>Task</th>
<th>Communication Type</th>
<th>Speech Type</th>
<th>Context (Location / Theme)</th>
</tr>
</thead>
<tbody>
<tr>
<td>You're giving a party for 12 people but you don't have a bowl big enough for the salad. You decide to ask your neighbor Cindy if she has one. You run into Cindy just as she's coming home and unlocking her door. You and Cindy have greeted each other and made a little small talk, but you're in a bit of a hurry and need to ask her to borrow her large salad bowl. *</td>
<td>Trans. / Inter.</td>
<td>Request</td>
<td>N/A</td>
</tr>
<tr>
<td>Yesterday you borrowed a bowl from your neighbor, Cindy. You have washed it and walked next door to return it. When Cindy answers her door, you thank her and invite her over for tea. She asks you if she can come over later (once she's done her laundry), but you have a doctor's appointment in a couple of hours and need to find out how much longer her laundry will take her before answering the question.*</td>
<td>Trans. / Inter.</td>
<td>Negotiation</td>
<td>N/A</td>
</tr>
<tr>
<td>You have decided to go out to a movie and have walked down to your nearest [theatre name]. The machines that let you buy tickets are all out of order. So, you get in line and wait for your turn with the cashier. Once you get to the front of the line, she asks you &quot;how may I help you?&quot;</td>
<td>Trans.</td>
<td>Response</td>
<td>Movie Theatre</td>
</tr>
<tr>
<td>You have bought tickets to a movie that you have been wanting to see for quite some time and are waiting in line at the concession stand. You eventually get to the front and the clerk asks &quot;What would you like?&quot;</td>
<td>Trans.</td>
<td>Response</td>
<td>Movie Theatre</td>
</tr>
<tr>
<td>You have decided to go out to a movie with a friend and have walked down to your nearest [theatre name]. You both get in line and wait for your turn with the cashier. Once you get to the front of the line, she announces that the movie you have come to see has sold out. You turn to your friend to figure out what to do.</td>
<td>Inter.</td>
<td>Negotiation</td>
<td>Movie Theatre</td>
</tr>
<tr>
<td>Task</td>
<td>Communication Type</td>
<td>Speech Type</td>
<td>Context (Location / Theme)</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>You are out at a new restaurant with some of your friends. The waiter approaches and asks if you are ready to order. One of your friends says yes. After the waiter collects the orders he turns to you and asks &quot;You ordered the hamburger, right?&quot;, but that isn't what you ordered, you ordered a cheeseburger.</td>
<td>Trans.</td>
<td>Negotiation</td>
<td>Local Eatery</td>
</tr>
<tr>
<td>You are out at a new restaurant with some of your friends. The waiter approaches and asks if you are ready to order. One of your friends says yes. After the waiter collects the orders he turns to you and asks &quot;You ordered the steak, right? What kind of potato you would like with that?&quot; You did order the steak but would like to know what potato options are available.</td>
<td>Trans.</td>
<td>Response / Request</td>
<td>Local Eatery</td>
</tr>
<tr>
<td>You are out at a new restaurant with some of your friends. The waiter approaches and asks what you would like to drink.</td>
<td>Trans.</td>
<td>Response</td>
<td>Local Eatery</td>
</tr>
<tr>
<td>You and your friend need some help with a vocabulary item during health class. Your friend has asked you to ask the teacher a question about a word (ailment) that she doesn't understand.</td>
<td>Trans. / Inter.</td>
<td>Response / Request</td>
<td>Illness</td>
</tr>
<tr>
<td>You are a secretary in a tile factory and you need two days off because your mother is ill. You approach your boss and ask if he has a few minutes. He says that he does and invites you into his office.</td>
<td>Trans. / Inter.</td>
<td>Request</td>
<td>Illness</td>
</tr>
<tr>
<td>You are at work and suddenly feel very ill. You run to the washroom and throw-up in the toilet. You clean yourself up and go back out to your boss to ask to go home.</td>
<td>Trans.</td>
<td>Request</td>
<td>Illness</td>
</tr>
<tr>
<td>You are shopping for a shirt in your favourite store and a sales person that you have not previously spoken with approaches you and asks: “How may I help you?”</td>
<td>Trans.</td>
<td>Response / Request</td>
<td>Shopping</td>
</tr>
<tr>
<td>You are browsing through a shop looking for a mother's day gift. You've found something that you think would be perfect, but you can't figure out the price of it. So, you walk over to the clerk to find out how much it is.</td>
<td>Trans.</td>
<td>Request</td>
<td>Shopping</td>
</tr>
<tr>
<td>The air mattress that you bought last week for some unexpected houseguests to sleep on seems to be leaking air through its valve. So, you are taking it back to the store. When you get to the returns counter, the clerk asks &quot;if you would like to exchange the air mattress or if you would like a refund&quot;.</td>
<td>Trans.</td>
<td>Response</td>
<td>Shopping</td>
</tr>
</tbody>
</table>

* Training Tasks

Trans. = Transactional; Inter. = Interactional
Appendix G – VocabNomad Entity-Relationship Diagram

The database entity-relationship diagram for the server is displayed below. It is subdivided into three sections so that the diagram details can be seen (Figure G.1, Figure G.2, and Figure G.3).

Figure G.1 The top portion of the entity-relationship diagram.

Figure G.2 The centre portion of the entity-relationship diagram.
Figure G.3 The bottom portion of the entity-relationship diagram.
Appendix H – Adaptive Testing Development

H.1 Test Development

The words for the adaptive test when it does not include system usage information are a combination of those provided through two different word frequency lists (Davies & Gardner, 2010; Leech et al., 2001). These lists were combined by appending one list to the other. Duplicate items were then removed. I also removed any potentially inflammatory content or content that might be considered inappropriate in a school setting (e.g., alcohol and sex).

H.2 Scoring

The adaptive and personalized tests were scored, by hand, using the guidelines described in Table H.1.

An example of a shallow response comes from ESM11: I bough[t] a new truck yesterday.

An example of a specific response for tiny from ESM08: The tiny molecule passes through the gap junction between cell.

An example of a sentence where the meaning was absent comes from ESM01: What does “sack” mean?

ESM11 also provided an example of an incorrect response: They’ve frozen the popsicles by mistake.

Students were still awarded the score if they transformed the target word and used it correctly as ESM04 did for mammal: [non-decipherable], dogs and pigs are types of mammals.

<table>
<thead>
<tr>
<th>Points</th>
<th>Part of Speech Use</th>
<th>Semantics of Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Incorrect</td>
<td>Incorrect/Absent</td>
</tr>
<tr>
<td>1</td>
<td>Correct</td>
<td>Incorrect/Absent</td>
</tr>
<tr>
<td>2</td>
<td>Incorrect</td>
<td>Shallow</td>
</tr>
<tr>
<td>3</td>
<td>Correct</td>
<td>Shallow</td>
</tr>
<tr>
<td>4</td>
<td>Incorrect</td>
<td>Specific</td>
</tr>
<tr>
<td>5</td>
<td>Correct</td>
<td>Specific</td>
</tr>
</tbody>
</table>
I faced challenges when scoring items that related to food pairings because of cultural assumptions that I held about which foods are typically paired or when they are eaten. So, I searched recipes from participants’ home cultures to see if the pairings that they used were real or if participants had just realized that things were foods and grouped them together.

Some participant responses also required knowledge of the participants to assess if the words were being used correctly. For example, ESM05’s response of “I living in west side of Toronto” for west indicates that the person understood that west is a direction or area but knowing that the person lives near Kipling station provides evidence that ESM05 has a more precise understanding of what west means.
Appendix I – Testing Instruments

A copy of the items from the morphology-based assessments are included. The PPVT-4 items are not included to protect the integrity of the test. The final exam questions that were used for the English Intensive 2 (EI2) class in Japan have also been included.

I.1 Morpho-orthographic Choice Task

Instructions:

In each pair of words, please circle the word that is spelled correctly.

<table>
<thead>
<tr>
<th>Word pairs:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Performance</td>
<td>Performance</td>
<td>32. Transmission</td>
</tr>
<tr>
<td>2. Fearful</td>
<td>Fearfull</td>
<td>33. Typical</td>
</tr>
<tr>
<td>3. Dayly</td>
<td>Daily</td>
<td>34. Complition</td>
</tr>
<tr>
<td>4. Fullness</td>
<td>Fullness</td>
<td>35. Natural</td>
</tr>
<tr>
<td>5. Reduction</td>
<td>Reduction</td>
<td></td>
</tr>
<tr>
<td>6. Shiney</td>
<td>Shiny</td>
<td></td>
</tr>
<tr>
<td>7. Relyable</td>
<td>Reliable</td>
<td></td>
</tr>
<tr>
<td>8. Tableing</td>
<td>Tabling</td>
<td></td>
</tr>
<tr>
<td>9. Description</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>10. Varyable</td>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>11. Glorious</td>
<td>Glorius</td>
<td></td>
</tr>
<tr>
<td>12. Producsion</td>
<td>Production</td>
<td></td>
</tr>
<tr>
<td>13. Running</td>
<td>Running</td>
<td></td>
</tr>
<tr>
<td>14. Swimmer</td>
<td>Swimer</td>
<td></td>
</tr>
<tr>
<td>15. Prunning</td>
<td>Pruning</td>
<td></td>
</tr>
<tr>
<td>16. Adventurous</td>
<td>Adventurous</td>
<td></td>
</tr>
<tr>
<td>17. Expansion</td>
<td>Expantion</td>
<td></td>
</tr>
<tr>
<td>18. Pityful</td>
<td>Pitiful</td>
<td></td>
</tr>
<tr>
<td>19. Mysterious</td>
<td>Mysterous</td>
<td></td>
</tr>
<tr>
<td>20. Trial</td>
<td>Treal</td>
<td></td>
</tr>
<tr>
<td>21. Diging</td>
<td>Digging</td>
<td></td>
</tr>
<tr>
<td>22. Introduction</td>
<td>Introducesion</td>
<td></td>
</tr>
<tr>
<td>23. Funny</td>
<td>Funy</td>
<td></td>
</tr>
<tr>
<td>24. Permission</td>
<td>Permission</td>
<td></td>
</tr>
<tr>
<td>25. Beautifully</td>
<td>Beautifully</td>
<td></td>
</tr>
<tr>
<td>26. Receiveable</td>
<td>Receivable</td>
<td></td>
</tr>
<tr>
<td>27. Normally</td>
<td>Normaly</td>
<td></td>
</tr>
<tr>
<td>28. Absorbtion</td>
<td>Absorption</td>
<td></td>
</tr>
<tr>
<td>29. Admireable</td>
<td>Admirable</td>
<td></td>
</tr>
<tr>
<td>30. Fullfil</td>
<td>Fulfill</td>
<td></td>
</tr>
<tr>
<td>31. Explanation</td>
<td>Explanasion</td>
<td></td>
</tr>
</tbody>
</table>
I.2 Morphological Structure Task

Training:
A. Farm. My uncle is a ____________________.
B. Help. My sister is always ____________________.

Test Items:
1. Warm. He chose the jacket for its ____________________.
2. Teach. He was a very good ____________________.
3. Permit. Father refused to give ____________________.
4. Profit. Selling lemonade in summers is ____________________.
5. Appear. He cared about his ____________________.
6. Express. ‘OK’ is a common ____________________.
7. Four. The cyclist came in ____________________.
8. Remark. The speed of the car was ____________________.
9. Protect. She wore glasses for ____________________.
10. Perform. Tonight is the last ____________________.
11. Expand. The company planned an ____________________.
12. Revise. The paper is his second ____________________.
13. Reason. Her argument was quite ____________________.
14. Major. He won the vote by a ____________________.
15. Deep. The lake was well known for its ____________________.
16. Equal. Boys and girls are treated with ____________________.
17. Long. They measured the ladder’s ____________________.
18. Adventure. The trip sounded ____________________.
19. Absorb. She chose the sponge for its ____________________.
20. Active. He was tired after so much ____________________.
21. Swim. She was a strong ____________________.
22. Human. The kind man was known for his ____________________.
23. Wash. Put the laundry in the ____________________.
24. Humor. The story was quite ____________________.
25. Assist. The teacher will give you ____________________.
26. Mystery. The dark glasses make the man look ____________________.
27. Produce. The play was a grand ____________________.
28. Glory. The view from the hill top was ____________________.

I.3 EI2 Final Exam Questions

Vocabulary
[ I ] Read each sentence and choose most appropriate answer for each blank.
1. This team won the championship for two ( ) years.
2. He ( ) by police officers, but they later realized that he was innocent.
3. My parents joined a ( ) farm last year and farm alongside others.
4. I had to accept my ( ).
5. He ( ) some troubles.
6. They broke up a month before, and now, they ( ).
7. Cancer is an (   ) illness for human beings.
8. Don’t get (   ) in the classroom!!!
9. We should develop (   ) to create a better world.
10. You can use my document as a (   ) for your paper.
11. This tennis team’s motto is “(   ) your victory”. Maybe, they have a lot of confidence.
12. He used failure as a (   ) to success.
13. She set the (   ) so that all of the slides had the same style.
14. She is (   ), among young women, as a fashion leader.
15. She (   ) money from the bank.
16. He was (   ) the many chances that he was given.
17. He is my swimming (   ).
18. He won 1st (   ) at the competition.
19. The teacher (   ) students to study hard.
20. The jacket is available in (   ) types.

Word Bank

a) prize / b) incurable / c) springboard / d) various / e) encountered / f) grateful for /
g) seize / h) withdrew / i) expects / j) cooperative / k) theme / l) naked / m) was arrested /

n) rival / o) fate / p) well-known / q) are getting back together / r) consecutive /
s) reference / t) a different point of view

[ II ] Choose the most appropriate definition for each word.
1. achievement, / 2. compete, / 3. courageous, / 4. delete, / 5. dominant,
6. controversy, / 7. evolutionary, / 8. establish, / 9. genius, / 10. infant,

Definitions

a) someone who is very smart or talented
b) relating to the way in which plants and animals develop and change increasingly over a long period of time; slow change
c) to remove or erase something
d) a serious argument about something that involves many people and continues for a long time
e) taking part in a race or sports event; trying to be the best at something
f) most common; more important, powerful, or commanding than others
g) the action of accomplishing something; something important that you succeed in doing by your own efforts
h) to start; to create; to stabilize
i) to be brave
j) a baby or very young child

[ III ] Read each definition and write an appropriate word. When the alphabet is written on the answer sheet, you must begin writing with that letter.
1. a great work of art
2. the activity or job that someone spends their time doing
3. strong feelings
4. the leader; an executive officer of a firm or corporation
5. to withdraw from action; to stop working
6. a secondary or explanatory title
7. to move something from one place to another
8. words of good sense and judgment, based especially on your experience of life
9. A paper or set of papers that are given to people who attend an event
10. a legal agreement between people, companies, etc

271
Rewrite these words into a sentence in the most appropriate way.
1. My [ from / my grandfather / are inherited / that my green eyes / family thinks ].
2. The [ the dinosaurs / a long time ago / of / happened / extinction ].
3. He [ my suggestion / and dismissed / just laughed / as unrealistic ].
4. They [ had broken / since / installed / the old one / a new window ].
5. She [ practice anymore / until she / rehearsed / her presentation / just couldn't ].

Listening
Watch the Guy Kawasaki presentation advice video and complete the notes below. You will see the video twice. There will be 30 seconds writing time in between.

<table>
<thead>
<tr>
<th>What</th>
<th>Guy Kawasaki always uses the 1) __________________________ format.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why</td>
<td>He uses it because 2) __________________________ suck.</td>
</tr>
<tr>
<td>Details</td>
<td>They suck because they are not 3) __________________________</td>
</tr>
<tr>
<td></td>
<td>are arrogant.</td>
</tr>
<tr>
<td></td>
<td>have people who 4) __________________________ believe an audience can 5) __________________________ from 100 feet away.</td>
</tr>
<tr>
<td></td>
<td>don’t have anything to say so they write more to back that up.</td>
</tr>
</tbody>
</table>

Researching and making a PowerPoint slide with notes
Using the text below, make two PowerPoint slides and save them in the 提出箱 with your full name, e.g. Daita Watanabe.
- You do not need to use all the information.
- Decide the topic of your slides and focus on the information you need.
- Include any important notes. (Do not write a script.)
- Do not use the Internet except to get pictures!

Pele

Pele is regarded by many players, critics and fans as the best soccer player ever. He was born in 1940 and started playing club football at 15. He took part in his first World Cup when he was 16 and went on to win three World Cups. He is in the Guinness World Records for scoring the most goals in a career. He actually scored 1281 goals in 1363 games. He was famous for being a very exciting player and scoring amazing goals.

Early years

His name is Edson Arantes do Nascimento. Interestingly, he was first called Edison, after the famous inventor Thomas Edison. However, his parents decided to remove the “I”. Unfortunately, when he was born, the doctor made a spelling mistake, so sometimes his name is still spelled Edison! His nickname is also interesting. He started being called Pele when he was a school student, but no one is sure where the nickname came from. Some think it was given to him by a priest, and others think his school friends gave it to him. Also, in his native language, Portuguese, it has no meaning!

Club career

272
Pele joined the Brazilian team Santos, in Sao Paulo, in 1956 and played his debut senior match at the age of 16. In that match, a 7-1 win, he scored the first of his many, many goals. He played for Santos until 1967 and during that time he won over twenty league titles and cups.

“I arrived hoping to stop a great man, but I went away convinced I had been undone by someone who was not born on the same planet as the rest of us” — Benfica goalkeeper Costa Pereira following the loss to Santos in 1962.

After 19 seasons with Santos, Pele retired but came out of retirement two years later to play for the North American Soccer League team New York Cosmos. Pele’s involvement helped to make soccer more popular in the United States. He ended three seasons with Cosmos by winning the league and played his final soccer match on October 1st 1977. The match was between Santos and Cosmos with Pele playing for Cosmos in the first half, and scored a goal. Then he played for Santos in the second. The game finished with a 2-1 win for Cosmos.

**World Cups**

1958

Pele was injured before this World Cup but his teammates insisted he should be selected. As a result, he didn’t play until the third match of the first round. In that match he gave an assist to the player Vava, then in the next match he scored the only goal. That goal made him the youngest player to score in a World Cup (he was just 17). Next, in the semi-final against France he scored a hat-trick and became the youngest player to do that in a World Cup, too. On June 19th 1958, he became the youngest player to play in a World Cup final, and he scored two goals as Brazil beat Sweden 5-2 and finished the tournament with six goals in four games.

1962

Many people expected Pele to be the star of this World Cup but he was injured in the first match. Brazil went on to win the competition, but Pele played no further part.

1966

This World Cup was famous for the terrible fouling on Pele by Bulgarian and Portuguese players. Pele was fouled many times and missed the second game because of injury. He returned for the third game but was injured because of a very bad foul. In those days, there were no substitutions, so the injured Pele had to stay on the pitch. Brazil were eliminated from the competition after the group stage, and after watching the video of the third match, Pele said he would never play in a World Cup again.

1970

Pele undid his decision to retire from World Cup soccer and went to the 1970 World Cup in Mexico. The Brazilian team contained so many stars that it is considered the most talent team ever. It played with five forwards and Pele’s role was of central importance. Throughout the tournament, Pele scored and created goals. In the final, he scored the first goal and created two others as Brazil beat Italy 4-1. After the match, the Italian Burgnich, who had to defend against Pele was quoted as having said the following:

“I told myself before the game, he’s made of skin and bones just like everyone else — but I was wrong”.

**Legacy**

Pele is considered to be the greatest player ever by most fans, critics and players, although younger fans may choose others because they have not seen his plays. It is perhaps most important that so many famous players say he is the best. Johan Cruyff, Bobby Moore, Zico, Cristiano Ronaldo and many more all say that no other
player was as talented as Pele. Maybe the best quote given about his level came during the 1970 World Cup when a British television commentator asked, “How do you spell Pele?”, with the response; “Easy: G-O-D.” Also, Pele himself was not so modest about his fame. One news reporter asked him if he was as famous as Jesus Christ and Pele answered by saying, "There are parts of the world where Jesus Christ is not so well known."

**Personal life**

Pele’s personal life has had several controversies. He has been married twice but has had several girlfriends. He has had several children with his various wives and girlfriends. In addition to those difficulties, in 2014 his son was arrested and sent to jail for 33 years because of his involvement with drugs.

**After football**

Since his retirement, Pele has been very busy. He has worked for the United Nations to help to protect the environment. He has been involved in football, television, writing and musicals. He famously appeared in the World War II movie *Escape to Victory* with Michael Caine and Sylvester Stallone. He has also been involved in advertisements for medicines and fashion, and has featured in several computer games.

Source: http://en.wikipedia.org/wiki/Pele
Appendix J – Japan: Course Worksheets & Reports

J.1 Course Worksheets

Examples of a student-completed worksheet can be seen in Figure J.1 and Figure J.2.

The content of one of the worksheets that was meant to support student discussion is included below. Students were given this worksheet to support the group-based discussions that they would be performing. It contained material that they had learned in a previous course

**Discussion Phrases**

**Asking for an opinion**
- How about you?
- What do you think (about this topic)?
- How do you feel about ...?
- What do you think about ...?
- What's your view on ...?
- What's your position (on this topic)?

**Giving an opinion**
- In my opinion, ...
- From my point of view, ...
- Personally, I think ...
- I personally feel that ...
- I feel that ...
- As far as I'm concerned, ...
- I think ...
- I'm certain that ...
- I firmly believe that ...

**Expressing agreement**
- I agree (with you).
- I think so too.
- I have the same opinion.
- I think you are right.
- I completely agree.
- I totally agree with you.
- Exactly
- Well, yes, you're absolutely right.

**Expressing disagreement**
- I disagree (with you).
- I totally disagree (with you).
- I don’t agree with you at all.
- I don’t completely agree with you (on that).
- I have a different opinion.
- I am afraid I have to disagree (with you).
- I’m sorry, but (I think) you’re mistaken
- I am sorry to disagree with you, but personally I think ...
- I can see your point of view, but ...
- The reason I feel this way is ...
- I feel this way because ...
- I believe this because ...

**Asking to talk**
- May I interrupt you for a moment?
- Sorry to interrupt you, but ...
- Excuse me, but I would like to say something.

**Asking for clarification**
- Excuse me, but could you say that again?
- I'm sorry, but I didn’t quite understand (what you said).
- Could you explain, please?
- I am afraid I don’t understand.
- Could you rephrase that?
## English Intensive II
### Vocabulary for PowerPoint

Please find the words that you don’t already know in VocabNomad and try to figure out what they are in Japanese. You should ONLY check in the dictionary to see if you are correct AFTER you have translated all of the words.

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>PowerPoint</td>
<td>denshi gazou, tsueltki</td>
</tr>
<tr>
<td>slide</td>
<td>itinen, gazou</td>
</tr>
<tr>
<td>outline</td>
<td>gairyaku</td>
</tr>
<tr>
<td>orientation</td>
<td>ichi</td>
</tr>
<tr>
<td>styles</td>
<td>jyoutai</td>
</tr>
<tr>
<td>animations</td>
<td>anime</td>
</tr>
<tr>
<td>transition</td>
<td>kirikae</td>
</tr>
<tr>
<td>slide show</td>
<td>tenpugazou</td>
</tr>
<tr>
<td>beginning</td>
<td>saisho, hajimari</td>
</tr>
<tr>
<td>current slide</td>
<td>yoyaku</td>
</tr>
<tr>
<td>rehearse</td>
<td>rensuu-suru,</td>
</tr>
<tr>
<td></td>
<td>rehersual-suru</td>
</tr>
<tr>
<td>timings</td>
<td>seigenjikan</td>
</tr>
<tr>
<td>presenter</td>
<td>happyosha, kaisetsusyou</td>
</tr>
<tr>
<td>slide sorter</td>
<td>gazoichiran</td>
</tr>
<tr>
<td>handout</td>
<td>purino</td>
</tr>
<tr>
<td>notes</td>
<td>kirekuchyo</td>
</tr>
<tr>
<td>title</td>
<td>shyudai</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>subtitle</td>
<td>fukudai</td>
</tr>
<tr>
<td>bullet points</td>
<td>youfien</td>
</tr>
<tr>
<td>bullet</td>
<td>jyuminyou</td>
</tr>
<tr>
<td>video</td>
<td>eizo</td>
</tr>
<tr>
<td>left aligned</td>
<td></td>
</tr>
<tr>
<td>right aligned</td>
<td></td>
</tr>
<tr>
<td>centered</td>
<td></td>
</tr>
<tr>
<td>justified</td>
<td>seitousuka-suru</td>
</tr>
<tr>
<td>footer</td>
<td>hosoku</td>
</tr>
<tr>
<td>duplicate slide</td>
<td>fukusesu - gazo</td>
</tr>
<tr>
<td>cut</td>
<td>kiriteru</td>
</tr>
<tr>
<td>copy</td>
<td>shymuraru</td>
</tr>
<tr>
<td>paste</td>
<td>haru, nerizuke-suru</td>
</tr>
<tr>
<td>reset</td>
<td>yarinaosu</td>
</tr>
<tr>
<td>hide</td>
<td>kakusu</td>
</tr>
<tr>
<td>new slide</td>
<td>shinki-gazou</td>
</tr>
<tr>
<td>theme</td>
<td>teima, shudai</td>
</tr>
</tbody>
</table>

Figure J.1 One of the worksheets that was completed by Nihon_S37.
### English Intensive II

#### Vocabulary for Mixed Martial Arts (MMA)

Some of the words have not been translated. Please find them in VocabNomad and try to figure out what they are in Japanese. You should ONLY check in the dictionary to see if you are correct AFTER you have translated all of the words.

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>legend</td>
<td>densetsu</td>
</tr>
<tr>
<td>mixed martial arts (MMA)</td>
<td>sogo-kakutogi</td>
</tr>
<tr>
<td>martial arts</td>
<td>bujyutsu</td>
</tr>
<tr>
<td>achievement</td>
<td>tassei</td>
</tr>
<tr>
<td>controversy</td>
<td>ronso, giron</td>
</tr>
<tr>
<td>words of wisdom</td>
<td>kakugen, meigen</td>
</tr>
<tr>
<td>infant</td>
<td>taiji</td>
</tr>
<tr>
<td>compete</td>
<td>kyoso-suru, arasou</td>
</tr>
<tr>
<td>well-known</td>
<td>yumeina</td>
</tr>
<tr>
<td>springboard</td>
<td>kikkake</td>
</tr>
<tr>
<td>myth</td>
<td>shinwa</td>
</tr>
<tr>
<td>illusion</td>
<td>genso, sakkaku</td>
</tr>
<tr>
<td>submission</td>
<td>kansetsuwaza</td>
</tr>
<tr>
<td>rear naked choke</td>
<td>hadaka-jime</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>choke</td>
<td>chissoku-jotai</td>
</tr>
<tr>
<td>pupils</td>
<td>deshi, seito</td>
</tr>
<tr>
<td>founder</td>
<td>sousetsusha</td>
</tr>
<tr>
<td>to be naturalized</td>
<td>kika-suru</td>
</tr>
<tr>
<td>quote</td>
<td>innyosuru</td>
</tr>
<tr>
<td>evolutionary</td>
<td>shinkano</td>
</tr>
<tr>
<td>point of view</td>
<td>mono no mikata</td>
</tr>
<tr>
<td>meditate</td>
<td>meiso-suru</td>
</tr>
<tr>
<td>pray</td>
<td>inoru</td>
</tr>
<tr>
<td>to be thankful</td>
<td>kanshasuru</td>
</tr>
<tr>
<td>to be grateful</td>
<td>keuryosuru</td>
</tr>
<tr>
<td>at peace</td>
<td>yasurakani, odayakani</td>
</tr>
<tr>
<td>reference</td>
<td>sanko bunken</td>
</tr>
<tr>
<td>chokehold</td>
<td>futan, shimekiri</td>
</tr>
</tbody>
</table>

---

**Figure J.2 One of the worksheets that was completed by Nihon_S37.**
J.2 Student Reports

Each student received a different combination of worksheets that were organized by pack (Table J.1).


The base pack contained information and activities that related to the pronunciation of long and short vowels, the pronunciation of vowel pairs, prefixes and their meanings, irregular plural nouns, the use of visual cues from images to aid comprehension, suffix meaning, finding root words, using context to disambiguate homographs, homophones, comparative adverbs, and proper adjectives.

Pack 1 contained exercises that required students to use context to determine word meanings, create the plural form of regular nouns, add the correct prefix to words, work with suffix meanings, find root words, use antonyms as context clues, put regular verbs into the past tense, and add suffixes that require some modification of the original word.

Pack 2 contains exercises that are related to creating the past-tense of regular verbs, using prefixes that have similar meanings, adding suffixes to words, creating the irregular plural possessive form of nouns, prefix meanings, working with antonyms, transforming adjectives into adverbs, subject-verb agreement, using context to support comprehension, comparative and superlative adverbs, suffix meaning, and homophones.

Pack 3 contains exercises about suffix meanings, transforming adjectives into adverbs, working with prefixes and suffixes, joining words, antonyms, determining meaning from context, identifying root words, and reading comprehension.

Pack 4 contains activities for working with suffixes, homophones, synonyms and antonyms, comparatives and superlatives, collective nouns, prepositions, (un)countable nouns, interrogative pronouns, and determining meaning from context.
Pack 5 contains worksheets that ask the learner to determine meaning from context, use appositives, idioms, comparatives and superlatives, prepositions, verb conjugation, and word decomposition.

Pack 6 contains activities that require word decomposition, idioms, determining meaning from context, conjugating verbs, plural possessive nouns, and reading comprehension.

Table J.1 The packs that were included in each students’ booklet.

<table>
<thead>
<tr>
<th>Username</th>
<th>Base Pack</th>
<th>Pack 1</th>
<th>Pack 2</th>
<th>Pack 3</th>
<th>Pack 4</th>
<th>Pack 5</th>
<th>Pack 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nihon_C01</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C02</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C03</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C04</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C05</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Nihon_C06</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_C07</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_C08</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C09</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C10</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C11</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C12</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_C13</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S01</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S02</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S03</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S04</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S05</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S06</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S07</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S08</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S09</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S10</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S11</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S12</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S13</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S14</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S15</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S16</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S17</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S18</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Username</td>
<td>Base Pack</td>
<td>Pack 1</td>
<td>Pack 2</td>
<td>Pack 3</td>
<td>Pack 4</td>
<td>Pack 5</td>
<td>Pack 6</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Nihon_S19</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S20</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S21</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S22</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S24</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S25</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Nihon_S26</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S27</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S28</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S29</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S30</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S31</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S32</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S33</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S34</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S35</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S36</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S37</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S38</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S39</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S40</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S41</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S42</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S43</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S44</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S45</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S46</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S47</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S49</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

In addition to these workbook packs, students received a report about their test performance and language abilities. The below content shows the template for one of the student reports. This content was included in a booklet that also contained the exercises that had been chosen based on each student’s test results and answers.

**About the Tests**

60 students took the tests: 1 from EI 1, 47 from EI 2, and 12 from other year 3 courses.
We ONLY tested your ability to recognize words, your ability to read words, and your ability to write words within sentences. None of the tests related to your ability to speak or listen in English but they do measure many of the skills and knowledge that are needed to learn English.

A native speaker of English is unlikely to get full marks on all of the tests.

**Overall**

This is an average of your score for all 3 types of tests.

It is out of a maximum of 100. Your average score was ______.

![Figure 1. Score range. Your score is marked with a coloured X.](image)

**Vocabulary (Word) Knowledge**

This was the picture test that we did many times. It can be used for native and non-native speakers of English to track changes in their vocabulary knowledge throughout their schooling.

The maximum possible score that a student could receive on this test was 76. Your average score was ______.

![Figure 2. Score range. Your score is marked with a coloured X.](image)
Recognizing the Correct Word

This was the test where you circled the correct word. Some of you called it a spelling test.

The maximum possible score that a student could get on this test was 35. You received a

![Figure 3. Score range. Your score is marked with a coloured X.](image)

Writing the Correct Word

This was the test where you had to change the word and put it in a sentence.

This test is the only test where the students from EI2 performed better than the other students. However, some of the other students earned similar scores to students in EI2.

The maximum possible score that a student could get on this test was 56. You received a

![Figure 4. Score range. Your score is marked with a coloured X.](image)

Other Feedback

282
J.3 School Report

**VocabNomad Project - Results**

**About the Tests**

60 students took the tests:
- 1 from EI 1
- 47 from EI 2
- 12 from other year 3 courses

We ONLY tested students’ ability to recognize words, their ability to read words, and their ability to write words within sentences. None of the tests related to their ability to speak or listen in English but they do measure many of the skills and knowledge that are needed to learn English.

It is worth noting that a native speaker of English is unlikely to get full marks on all of the tests. However, they were used to test the full range of student abilities. We would not have been able to trust the test results if everyone had achieved high scores.

**Overall**

This is an average of student scores for all 3 types of tests (Figure 1).

**Vocabulary (Word) Knowledge**

We used the PPVT-4. It asks students to look at 4 pictures and select the one that matches the

![Boxplot Diagram](image)

*Figure 1 The average of student scores (%) on the tests that were administered as part of the study. This shows a fairly wide range in student abilities.*
Figure 2. Student scores on the general test of vocabulary knowledge.

word that I read aloud. It can be used for native and non-native speakers of English to track changes in their vocabulary knowledge throughout their schooling. This is a difficult test where students select the picture that matches the word.

This test was performed 3 times. Students’ average performance across all three tests is shown in Figure 2. It is worth noting that student scores improved around 12% between the first test and second test but did not change more than a few percent between the second and third test.

Recognizing the Correct Word

This was the test where students circled the correct word. They called it a spelling test. Most students performed well on this test (Figure 3).

Example question: skateing skating

This gives us information about students’ spelling skills and their ability to recognize the parts of words that are used to change their meaning or use within a sentence.

Figure 3. Student score range on their Recognizing the Correct Word Test.
Writing the Correct Word

This test asked students to change the word and put it in a sentence. Student performance on this test varied a lot (Figure 4).

Example: Farm. My uncle is a ____________.

This test is the only test where the students from EI2 performed better than the other students. However, some of the other students earned similar scores to students in EI2.

![Figure 4. Student score range on the Writing the Correct Word Test.](image)

What These Tests Tell Us

Students were stronger at recognizing which word was spelled correctly than they were at producing correctly spelled words when they were asked to use them in a sentence. An analysis of the mistakes that they made on both tests indicates that they could improve their understanding of prefixes and suffixes. They could also improve their ability to manipulate words by adding prefixes or suffixes. Many of the activities that were given to students at the end of the study focused on these skills because they contribute to students’ ability to write, understand new words when they see them, and understand texts when they are reading them.

Students had trouble with the following when adding a suffix to a word

- Knowing when the final vowel of a word should be removed (e.g., table -> tabling)
- Knowing when the final consonant should be doubled (e.g., run -> running or prune -> pruning)
- Knowing when y should become i (e.g., day -> daily)
Other Measures

Vocabulary Portion of the Final Exam

The vocabulary portion of the final exam shows that most students learned the words that were presented through VocabNomad. Three quarters of the students scored 70 percent or higher on this portion of the exam.

Self-Assessed English Language Ability

Students were asked to rate their ability to read, understand, and speak in English. This was done at the beginning, in the middle, and at the end of term. These ratings show that student self-perceptions of their English abilities increased when using VocabNomad. This small increase indicates greater confidence in their abilities.

The distribution of students’ average self-assessed English language abilities can be seen below (Figure 5).

![Figure 5 The average of student self-assessed English-language ability.](image)

VocabNomad Use

Students used VocabNomad fairly extensively during the first half of the term. This was when we had given them highly structured activities to perform using the application. This is also when we see an increase in their vocabulary knowledge and their self-assessed English language ability.
Some students continued to use VocabNomad during the second half of the term but many chose not to use it while preparing their presentations.

However, students successfully created vocabulary lists that were distributed to their classmates. These vocabulary lists were intended to teach their classmates the words that would be needed to understand the presentation topic that each group chose.

In addition to the above analysis, advanced analysis techniques from the area of data mining were applied. These found that there was a relationship between how students used VocabNomad and the improvement on their PPVT-4. This relationship was especially noticeable for the students who improved the most but rated their English abilities as being poor.

**What this means**

The ability of students to learn English through the English resources that were part of VocabNomad indicates that they are ready to start doing this. Adding opportunities to learn English by asking students to use English resources to learn the meaning of words could benefit them.

These opportunities can come from literature, games, mobile applications, or other resources. The important thing is that students should be given the opportunity to interact with English materials that they find difficult. However, they will need additional training in how to decompose words into their parts or use contextual clues to figure out what words mean. This is something that could start during their first year since learning more about suffixes and prefixes then could help them achieve better scores on tests of their reading, listening, and writing abilities.
Appendix K – Japan: Student Data

Table K.1 and Table K.2 summarize student exposure to the English language outside of their school environment. These tables include information on the types of English media that students use, the time that they have spent in English-language environments, and their parents’ knowledge of English. Details of each student group are provided in the Treatment and Control sub-sections. The time spent in English-language environments is additive across the table columns (Y – year(s), M – month(s), W – week(s), D – day(s)). For example, Nihon_C12 has spent 1 month and 2 days in English-language environments.

Table K.3 provides information about the types of device personalization that students performed. This includes their language settings which show implied preferences that may be partly based on the student’s comfort with English.

Table K.1 Control group exposure to English. Check marks indicate an exposure source. Question marks indicate that a participant did not answer the question.

<table>
<thead>
<tr>
<th>Student</th>
<th>Parental English</th>
<th>Time in English Environment</th>
<th>Types of English Media Used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mom</td>
<td>Dad</td>
<td>Y</td>
</tr>
<tr>
<td>Nihon_C01</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Nihon_C02</td>
<td>✓</td>
<td>✓</td>
<td>2</td>
</tr>
<tr>
<td>Nihon_C03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C04</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C07</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C08</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Nihon_C09</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Nihon_C10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_C11</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Nihon_C12</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Nihon_C13</td>
<td></td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>
### Table K.2 Treatment group exposure to English and class project group membership.

<table>
<thead>
<tr>
<th>Student</th>
<th>Group</th>
<th>Parental English</th>
<th>Time in English Environment</th>
<th>Types of English Media Used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mom</td>
<td>Dad</td>
<td>Y</td>
</tr>
<tr>
<td>Nihon_S01</td>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S02</td>
<td>5</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S03</td>
<td>10</td>
<td>✓</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Nihon_S04</td>
<td>16</td>
<td>✓</td>
<td>4</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S05</td>
<td>4</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S06</td>
<td>14</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S07</td>
<td>10</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S08</td>
<td>16</td>
<td>✓</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Nihon_S09</td>
<td>11</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S10</td>
<td>4</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S11</td>
<td>6</td>
<td>✓</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Nihon_S12</td>
<td>3</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S13</td>
<td>3</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S14</td>
<td>1</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S15</td>
<td>16</td>
<td>✓</td>
<td>✓</td>
<td>2</td>
</tr>
<tr>
<td>Nihon_S16</td>
<td>13</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S17</td>
<td>2</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S18</td>
<td>11</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S19</td>
<td>7</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S20</td>
<td>5</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S21</td>
<td>6</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Nihon_S22</td>
<td>11</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S23</td>
<td>8</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S24</td>
<td>13</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S25</td>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
</tr>
<tr>
<td>Nihon_S26</td>
<td>7</td>
<td>✓</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Nihon_S27</td>
<td>14</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S28</td>
<td>13</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S29</td>
<td>12</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S30</td>
<td>5</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S31</td>
<td>2</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S32</td>
<td>6</td>
<td>✓</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Nihon_S33</td>
<td>9</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S34</td>
<td>1</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S35</td>
<td>12</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S36</td>
<td>14</td>
<td>✓</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Nihon_S37</td>
<td>9</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nihon_S38</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Student</td>
<td>Parental English</td>
<td>Time in English Environment</td>
<td>Types of English Media Used</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
<td>----------------------------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td>Nihon_S39</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S40</td>
<td>40</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Nihon_S41</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S42</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S43</td>
<td>2</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Nihon_S44</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S45</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S46</td>
<td>15</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Nihon_S47</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nihon_S49</td>
<td>8</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table K.3 Student-performed device personalization. Check marks indicate that the feature was enabled.

<table>
<thead>
<tr>
<th>Student</th>
<th>Keyboard</th>
<th>Language</th>
<th>Pinyin</th>
<th>Voice Typing</th>
<th>Spell checker</th>
<th>Google Translate</th>
<th>No. Apps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nihon_S01</td>
<td>English (US)</td>
<td>English (US)</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>Nihon_S02</td>
<td>QWERTY</td>
<td>English (US)</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S03</td>
<td>English (UK)</td>
<td>English (UK)</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S04</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S05</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S06</td>
<td>English (US)</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S07</td>
<td>English (US)</td>
<td>English (US)</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Nihon_S08</td>
<td>English (US)</td>
<td>English (US)</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S09</td>
<td>English (US)</td>
<td>English (US)</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S10</td>
<td>English (US)</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Nihon_S11</td>
<td>English</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Nihon_S12</td>
<td>English (US)</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S13</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>Nihon_S14</td>
<td>English (US)</td>
<td>English (US)</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S15</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S16</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Nihon_S17</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S18</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S19</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S20</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S21</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>Nihon_S22</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S24</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S25</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Nihon_S26</td>
<td>English (UK)</td>
<td>English (UK)</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>Student</td>
<td>Keyboard</td>
<td>Language</td>
<td>Pinyin Typing</td>
<td>Spell checker</td>
<td>Google Translate</td>
<td>No. Apps</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Nihon_S27</td>
<td>English (UK)</td>
<td>English (UK)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S28</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S29</td>
<td>English (US)</td>
<td>English (Canada)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S30</td>
<td>English (US)</td>
<td>English (Canada)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S31</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Nihon_S32</td>
<td>English (US)</td>
<td>English (Canada)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S33</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S34</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S35</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Nihon_S36</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S37</td>
<td>English (US)</td>
<td>English (Canada)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S38</td>
<td>English (US)</td>
<td>English (Canada)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Nihon_S39</td>
<td>English (US)</td>
<td>English (Canada)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S40</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S41</td>
<td>Alphabet (QWERTY)</td>
<td>Accented English</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S42</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S43</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Nihon_S44</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S45</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S46</td>
<td>English (US)</td>
<td>English (Canada)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S47</td>
<td>English (US)</td>
<td>English (US)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Nihon_S72</td>
<td>iWnn IME - Jap</td>
<td>Japanese</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

*Cleared all installed apps*
Appendix L – Publications Related to this Thesis

The relationship of the below publications to my thesis is detailed in Table L.1.

Peer Reviewed Book Chapters


Peer Reviewed Journal Articles


Peer Reviewed Conference Papers and Posters in Main Proceedings


Peer Reviewed Abstracts


Invited


* Presenter
<table>
<thead>
<tr>
<th>Publication</th>
<th>Related Thesis Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>Chapter 6</td>
</tr>
<tr>
<td>J1</td>
<td>Chapter 7 and Chapter 8</td>
</tr>
<tr>
<td>C4</td>
<td>Chapter 3 and Chapter 4</td>
</tr>
<tr>
<td>C3</td>
<td>Chapter 5</td>
</tr>
<tr>
<td>C2</td>
<td>Chapter 4</td>
</tr>
<tr>
<td>C1</td>
<td>Chapter 3 and Chapter 4</td>
</tr>
<tr>
<td>A3</td>
<td>Chapter 8</td>
</tr>
<tr>
<td>A2</td>
<td>Chapter 5 &amp; Chapter 6</td>
</tr>
<tr>
<td>A1</td>
<td>Chapter 3</td>
</tr>
<tr>
<td>I2</td>
<td>Chapter 3, Chapter 6, Chapter 7, and Chapter 8</td>
</tr>
<tr>
<td>I1</td>
<td>Chapter 5</td>
</tr>
</tbody>
</table>