

Maximizing feedback for language learning:
English language learners' attention, affect, cognition and usage of
computer-delivered feedback from an English language reading proficiency
assessment

by

Maggie Dunlop

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Curriculum, Teaching and Learning
Ontario Institute for Studies in Education / University of Toronto

© Copyright by Maggie Dunlop 2017

ProQuest Number:10272178

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10272178

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Maximizing feedback for language learning:

English language learners' attention, affect, cognition and usage of computer-delivered feedback from an English language reading proficiency assessment

Maggie Dunlop

Doctor of Philosophy

Department of Curriculum, Teaching and Learning

University of Toronto

2017

Abstract

Attention to personalised feedback for language learning is increasing as computer-based assessment increases practicality, but little attention has been paid to how language learners interact with and use feedback from computer-based assessments. The purposes of the present research were two-fold: to investigate how adult immigrant English language learners engaged with and processed computer-based feedback on their English reading skills, and to explore how these learners used feedback depending on their processing outcomes, psychological characteristics, and English proficiency. To examine these issues, six data sources were analysed using mixed methods for complementary and developmental purposes through interviews, surveys, language assessments, and eye tracking with 102 adult immigrant English language learners in Canada. Data were analysed using qualitative coding and analysis and quantitative methods such as regression analyses and latent class profiling. Results were organized and synthesized by research questions.

Study findings were that the personalised sections received most attention, particularly visual results, but detailed descriptive text was useful at intermediate stages of feedback processing and usage. Learners' cognitive and affective strategies for negotiating feedback included emotional reactions, deflecting responsibility for negative feedback, critically evaluating report content,

negotiating comprehension difficulties, and relating the report to their own lives. Learners were generally positive about personalised feedback, adapted it for their own purposes, and used known affective and cognitive strategies, confirming earlier research in these areas. In addition, confirming other previous research, major factors impacting understanding and usage were external circumstances such as English language environment and language proficiency. A mastery goal orientation, trust in report content, reflection on English skills, and desire to use the report, were positively associated with report usage.

Implications included an observed need to fully factor feedback design into test design where impact/effects/outcomes are a guiding principle in test validation processes. From an instructional perspective, a key implication was the need to embed feedback in a high-quality, regular, and social learning environment. Further research is required to understand how feedback design can be personalized to promote more constructive feedback usage in learners with different background characteristics.

Acknowledgements

04(9) forever

Primary thanks to the language learners I have worked with over the years, for illuminating to me the constellation of hopes, dreams, fears, skills and experiences that every one of us brings to learning a language. I hope you have experienced success on your journeys.

Many thanks to my academic supervisor, Dr. Eunice Eunhee Jang, for her guidance, direction, and, quite frankly, pushing, so that I could achieve much more than I originally imagined. Many thanks also to my committee members Dr. Alister Cumming and Dr. Jim Slotta, to my external examiner Dr. Ari Huhta, and my internal examiner Dr. Earl Woodruff, for their sound advice, careful readings of my dissertation, and excellent feedback.

Sincere thanks to the managers and teachers at the five LINC programs who introduced me to their English language learners, and to Paragon Testing Enterprises for their role in supporting this dissertation. Heartfelt thanks also to Garfield Wu, Yongfei Wu, and friends, for helping me recruit participants by word of mouth, and to Hyun Wook Jo, Joe Caprara and other Education Commons colleagues, who helped me out of programming and hardware dead ends.

I deeply appreciate the financial support that I received during my doctoral studies and for this dissertation. I acknowledge the support of OISE/University of Toronto Funding Grants and Academic Excellence Awards, Ontario Graduate Scholarships, a CELPIP-General Doctoral Studies Research Grant, a Paragon Research Award, and a TOEFL Small Grant for Doctoral Research in Second or Foreign Language Assessment.

Thanks to my colleagues at Cambridge English Language Assessment for their encouragement and moral support while I endured the difficult task of simultaneously working full time and writing up a dissertation.

Finally, thanks to my friends, family and housemates, who have carefully avoided asking how my dissertation was coming along for rather longer than many of us expected, and brought joy to my life when the going was tough.

Table of Contents

Acknowledgements.....	iv
Table of Contents.....	v
List of Tables.....	ix
List of Figures.....	xi
List of Appendices.....	xiii
Chapter 1: Introduction.....	1
1.1 Introduction.....	1
1.2 Feedback from second language tests for assessment-for-learning purposes.....	2
1.3 Attention and processing in language learning.....	4
1.4 Learner characteristics and interaction with language learning feedback.....	6
1.5 Purpose and significance of the study.....	8
1.6 Overview of dissertation.....	10
1.7 Terminology for the study.....	10
Chapter 2: Literature review.....	14
2.1 Introduction.....	14
2.2 The place of feedback in language learning.....	14
2.3 Uses of cognitively diagnostic assessment for language learning.....	19
2.4 Language learning feedback validity.....	22
2.5 Self-regulated learning in language learning and its relationship with feedback.....	26
2.5.1 Cognitive strategies while dealing with second language material.....	27
2.5.2 Affective strategies while dealing with language learning challenges.....	28
2.5.3 Interaction of cognition and affect.....	30
2.5.4 Monitoring cognition of feedback on second language learning.....	31
2.5.5 Summary.....	33
2.6 Individual characteristics that interact with language learning feedback usage.....	35
2.6.1 Second language learning environment, goals and motivations.....	35
2.6.2 Goal orientation & beliefs about intelligence.....	37
2.6.3 Language proficiency.....	40
2.6.4 Self-assessment of language proficiency.....	42

2.6.5	Summary	44
2.7	Feedback design recommendations.....	44
Chapter 3: Methods.....		47
3.1	Research design overview	48
3.2	Study context.....	49
3.3	Participants	51
3.3.1	Participant recruitment.....	51
3.3.2	Participant characteristics	54
3.4	Data collection procedures	55
3.5	Instruments.....	57
3.5.1	First meeting: Background survey	57
3.5.2	Second meeting: Assessments	57
3.5.3	Third meeting: Receiving the feedback report.....	60
3.5.4	Fourth meeting: Delayed recall interview protocol	67
3.6	Overview of data sources	67
3.7	Analyses	68
3.7.1	Preparatory analyses	70
3.7.2	Data analyses by research theme	92
3.8	Conclusion.....	96
Chapter 4: Results.....		97
4.1	Theme 1 – Individual characteristics	97
4.1.1	RQ 1.1: What are the occupational contexts of adult immigrant English language learners in Canada?.....	97
4.1.2	RQ 1.2: What are the English language environments of adult English immigrant language learners in Canada?	99
4.1.3	RQ 1.3: What are the English language goals of adult immigrant English language learners in Canada?.....	101
4.1.4	RQ 1.4: What are typical goal orientation profiles and beliefs about intelligence among adult immigrant English language learners in Canada	102
4.1.5	RQ 1.5 What are typical English reading proficiency profiles among adult immigrant English language learners in Canada?	106
4.1.6	RQ 1.6: How do adult immigrant English language learners in Canada perceive their English reading proficiency?.....	109
4.1.7	RQ 1.7: What relationships are observable between individual characteristics? .	112

4.2	Theme 2 – Attention.....	115
4.2.1	RQ 2.1: To which aspects of feedback reports do language learners report paying attention?	115
4.2.2	RQ 2.2: To which aspects of feedback reports are language learners observed to pay attention?.....	119
4.2.3	RQ 2.3: Which aspects of feedback reports can be recalled by language learners one month later?	126
4.3	Theme 3 –Interaction with report content.....	128
4.3.1	RQ 3.1: What affective and cognitive strategies do language learners report when receiving a feedback report?.....	128
4.3.2	RQ 3.2: What processing outcomes do language learners report after receiving a feedback report?	135
4.3.3	RQ 3.3: In what ways are language learners’ processing outcomes related?	138
4.4	Theme 4 – Usage.....	142
4.4.1	RQ 4.1 How do language learners use a feedback report to plan their learning?..	142
4.4.2	RQ 4.2 What are the relationships between individual characteristics, attention and processing experience, and usage of feedback?	153
Chapter 5:	Discussion.....	163
5.1	Discussion 1: Relationship between reported and observed attention to feedback.....	164
5.2	Discussion 2: Processing of feedback report content.....	167
5.2.1	Cognitive and affective processes when receiving feedback on second language learning	167
5.2.2	Affective and cognitive outcomes after processing feedback on second language learning	170
5.3	Discussion 3: Factors that mediate usage of feedback on second language learning ..	173
5.3.1	Motivation and environment.....	174
5.3.2	Goal orientation and beliefs about intelligence	176
5.3.3	Language proficiency and self-assessment.....	178
5.3.4	Attention, cognitive processing and affective processing outcomes	180
Chapter 6:	Conclusions.....	184
6.1	Summary of results.....	184
6.2	Limitations of the study.....	187
6.3	Implications and suggestions	188
6.4	Areas for further research.....	190

References.....	192
Appendices.....	220
Appendix 1 Background survey.....	220
Appendix 2 Self-assessment tool.....	224
Appendix 3 Eye-tracking / interview protocol.....	225
Appendix 4 Report survey items.....	227
Appendix 5 Delayed recall interview protocol.....	229
Appendix 6 Exploratory factor analyses.....	230
Background survey: Goal orientation items.....	230
Report survey: Items for perceptions about amount of information.....	231
Report survey: Items for perceptions of content of report.....	232
Report survey: Items for planned usage of report.....	233

List of Tables

Table 1 Assessment and learning terms.....	11
Table 2 Language measurement terms	11
Table 3 Psychological terms	12
Table 4 Eye-tracking terms.....	13
Table 5 Total participants by data source	52
Table 6 Number of participants by participating school.....	53
Table 7 Number of participants by word-of-mouth network.....	54
Table 8 Reading subskills included on the report.....	59
Table 9 Report learning suggestions, by skill.....	63
Table 10 Overview of data sources.....	67
Table 11 Overview of preparatory data analysis	69
Table 12 Overview of theme-specific data analysis	69
Table 13 Spearman's rho correlations among goal orientation items.....	72
Table 14 List of items with codes.....	73
Table 15 Spearman's rho correlations among beliefs about intelligence items.....	74
Table 16 Pearson's r correlations between goal orientation and beliefs about intelligence composite variables.....	74
Table 17 Pearson's r correlations between estimates of skill mastery probabilities and total raw score	76
Table 18 Descriptive statistics for self-assessment scores, by skill.....	77
Table 19 Spearman's rho correlations between self-assessment items	78
Table 20 Evolution of stimulated recall interview codes: initial tentative codes	82
Table 21 Example coding using excerpts from two participants: initial tentative codes.....	82
Table 22 Evolution of stimulated recall interview codes: intermediate revised codes.....	84
Table 23 Evolution of stimulated recall interview codes: final codes with categories.....	84
Table 24 Example coding using excerpts from two participants: final codes with categories	85
Table 25 Participant interaction profiles: stimulated recall interviews	87
Table 26 Spearman's rho correlations among items assessing cognition and affect for amount of information.....	89

Table 27 Spearman's rho correlations among items assessing cognition and affect for report content.....	90
Table 28 Spearman's rho correlations among intended use items.....	91
Table 29 Phi coefficient correlations among discussion topics (clustered where possible).....	91
Table 30 Participant goal orientations: latent class profile models comparison table.....	104
Table 31 Probabilities of skill mastery: latent class profile models comparison table.....	107
Table 32 Self-assessment discrepancy: latent class profile models comparison table.....	110
Table 33 Eye-tracking interview participant interaction profiles.....	129
Table 34 Participant processing: latent class profile models comparison table.....	138
Table 35 A comparison of report content that was observably used in delayed recall interview participants' written plans, and how it was used in participants' written plans.....	152
Table 36 Additional activities reported by delayed recall interview participants that were also mentioned in their report.....	153
Table 37 Significant relationships between initial report usage and background, attention and processing variables.....	174
Table 38 Summary of exploratory factor analysis results for goal orientation items.....	230
Table 39 Summary of exploratory factor analysis results for items on perceptions about amount of information: all survey items.....	231
Table 40 Summary of exploratory factor analysis results for items on perceptions about amount of information: used survey items.....	232
Table 41 Summary of exploratory factor analysis results for items on perceptions of content of report.....	232
Table 42 Summary of exploratory factor analysis results for items on planned usage of report.....	233

List of Figures

Figure 1. Data collection activities and order	55
Figure 2. Example of introduction to report (fictional learner)	61
Figure 3. Example of report skill descriptions and learner personal results (fictional learner).....	62
Figure 4. Example of report suggestions section (for skills VOC and INF)	63
Figure 5. Example of report planning section (first plan outline only)	65
Figure 6. The mean scores on each goal orientation, for each class.....	104
Figure 7. The mean probability of mastery on each English reading skill, for each class	108
Figure 8. Plot for test score vs. perceived score in this study.....	109
Figure 9. Relationship between total raw score and overall self-assessment score.....	110
Figure 10. Relationship between total raw score and overall self-assessment score, with latent profile classes.....	111
Figure 11. Percentage of participants reporting amount of time spent looking at each section of the report.	116
Figure 12. Eye-tracking traces for a participant with very low skill mastery.....	122
Figure 13. Eye-tracking traces for a participant with mixed skill mastery: BK	123
Figure 14. Eye-tracking traces for a participant with high skill mastery: BO	123
Figure 15. Eye-tracking traces showing uneven fixation distribution on skill titles	125
Figure 16. Hierarchy of recall for report sections.....	128
Figure 17. Box plots showing distribution of mean scores for each report processing construct	137
Figure 18. The mean scores on each variable, for each class	140
Figure 19. Distributions of which people want to talk to	148
Figure 20. Mean 'talk to classmates' score by wanting to discuss specific topics	149
Figure 21. Percentage of participants choosing each number of skills, by number of report sections paid a lot of attention to	154
Figure 22. Relative amount of attention noted in the relevant data sources: Sections with most attention listed first	164

Figure 23. Key and secondary processing outcomes, with processing strategies that support stronger outcomes	173
Figure 24. Key factors mediating how language learners use non-teacher-mediated feedback	186

List of Appendices

Appendix 1 Background survey.....	220
Appendix 2 Self-assessment tool.....	224
Appendix 3 Eye-tracking / interview protocol.....	225
Appendix 4 Report survey items.....	227
Appendix 5 Delayed recall interview protocol.....	229
Appendix 6 Exploratory factor analyses.....	230

Chapter 1: Introduction

1.1 Introduction

Educators conduct assessments for various purposes in today's educational systems, with attention particularly turning to accountability purposes in the last decade or so. However, accountability-focussed assessments are often accused of failing to provide information useful to personal learning and classroom instruction, such as how much a student can do on various relevant skills (cf. Aydeniz & Southerland, 2012; Maier, 2009; Parkison, 2009). Meanwhile in the classroom, effective diagnosis is a particular issue where teachers receive large numbers of new students annually, for example in university foreign language programs, or where a teacher is entirely absent, for example where a learner is engaging in self-directed study. However, recent developments in technology are making such classroom diagnosis and feedback more common and possible, and methodologies are emerging (cf. Dikli & Bleyle, 2014; El Ebyary & Windeatt, 2010; Lee, Cheung, Wong, & Lee, 2013; Murphy, 2010; Jang, 2005). However to date, while much discussion around second language learning feedback has addressed the impact of feedback formats (Ellis, Loewen, & Erlam, 2006; Ferris, 2010; Li, 2014; Murphy & Roca de Larios, 2010; Nassaji, 2016; Opitz, Ferdinand, & Mecklinger, 2011), little research has been carried out that investigates the different ways in which second language learners interact with and use feedback from second language assessments (Jang, 2005; Fernandez-Toro & Hurd, 2014; Jang, Dunlop, Park, & van der Boom, 2015; Wagner, 2015). It is to this gap in the literature that this study contributes.

There were two main purposes of the research. The research studied how adult immigrant English language learners engage with and process feedback on their English language reading skills when delivered by computer. The research also studied how adult English language learners use feedback differently depending on how their processing outcomes, individual psychological characteristics, and English language proficiency. The ultimate aim was to identify how these factors interacted, and therefore contribute to understanding of second language learner engagement with computer-based feedback formats, particularly in the diagnostic context. The study builds on and contributes to the literature in the fields of second language assessment, assessment-for-learning, and educational technology, each of which I touch upon below, and in greater depth in the literature review.

1.2 Feedback from second language tests for assessment-for-learning purposes

Educators often seek to observe students' skills in order to provide guidance and instruction terms of learning and personal development. Due to Black and Wiliam's (1998) work followed by Broadfoot et al.'s (1999) work, awareness of the importance of structured observation for instructional purposes has developed into a well-researched field of classroom assessment (Hattie & Timperley, 2007; Parr & Timperley, 2010; Shute, 2008), centred around terminology such as formative assessment and assessment-for-learning. However, a standardized means for deriving information for assessment-for-learning has not generally been a focus of research in second language education (Burner, 2014; Ke, 2006; Ketabi & Ketabi, 2014), although the field of early years reading (in either first or second language) has devoted substantial effort to this issue (Bartlett, Dowd, & Jonason, 2015; Dubeck & Gove, 2015; Hill 2011; Snow, 2014; Snow, Burns, & Griffin, 1998).

Simultaneously, the field of second language assessment has developed into an established field that generally emphasizes the development of high quality measures of second language ability (cf. leading field journals: *Language Assessment Quarterly*; *Language Testing*). Given the socially and cognitively multifaceted nature of language ability, as well as, from a psychometric perspective, its status as a latent trait, a variety of validation processes were established as a means of building arguments supporting claims made about language tests in terms of defensibility (Bachman, 2005; Fulcher & Davidson, 2007; Kane, 1992; Mislavy & Haertel, 2006; Weir, 2005). However, none of the validation or development frameworks have devoted substantive attention to how results of second language tests need to be communicated to learners or embedded in programs of learning, although research in dynamic assessment works on bridging the issue (Lantolf & Poehner, 2004; Leung, 2007; Poehner, 2007).

It can be argued that historically one of the reasons for this disconnect was simply an issue of logistics and technology. The computing power necessary to develop high quality measures of second language ability – a notoriously difficult ability to assess – that can deliver sufficiently detailed information for use in learning, has only recently been developed (Musso & Cascellar, 2009). Likewise, the ability of learners and teachers to access externally developed and validated tests and assessments that can inform second language learning was previously limited by logistical challenges such as the necessity of fixed testing dates and locations for externally

developed testing material, and by time limitations in terms of interpreting information for learners and teachers to use (ETS, 2011; Hawkey & Milanovic, 2013; Sasaki, 2008).

However, recent technological developments make realistic previously impossible assessment, learning and classroom practices. These new horizons primarily result from the rise of computer-based assessment and reporting (ETS, 2011; van der Kleij, Eggen, Timmers, & Veldkamp, 2012). In particular, new horizons are opening in terms of creating authentic tasks, combining audio and visual task stimuli (Wagner, 2008), creating new response formats, increasing efficiency and time burden of delivery, adapting assessments to learner expertise (Chang, 2015; Deville & Chalhoub-Deville, 1999), and assessing multiple skills simultaneously (DiBello, Stout, & Roussos, 1995; Leighton & Gierl, 2007; Yan, Almond, & Mislevy, 2004). These technological developments offer an excellent opportunity to increase access to language assessments that have undergone rigorous validation processes, and the present study uses results from an emerging development in assessment: cognitive diagnostic modelling (DiBello, Stout, & Roussos, 1995; Jang, 2005; Kim, 2015; McGlohen & Chang, 2008; Nichols, 1994), to be discussed further in the next chapter's literature review.

In addition, one of the great areas of potential in computer-based assessment is in delivery of results, because computer-based assessment potentially resolves many of the logistical assessment and feedback challenges previously experienced by learners and teachers. In particular, computer-based assessment offers the ability to adjust feedback timing and tailor feedback content to an extent and on a scale that was previously impractical (Corbalan, Paas, & Cuypers, 2010; El Ebyary & Windeatt, 2010; Lee, Cheung, Wong, & Lee, 2013). Taking advantage of computing power and new technology therefore offers the opportunity to bring together the fields of second language assessment and second language learning in ways not previously logistically possible.

However as with previous work on classroom-based assessment for learning, the bulk of attention on feedback in literature for computer-based assessment for learning goes to learners' experiences and learning outcomes from online learning environments (cf. Hwang, Shih, Ma, Shadiev, & Chen, 2016; Peterson, 2016), impact of feedback formats on learning outcomes (cf. El Ebyary & Windeatt, 2010; Lee, Cheung, Wong, & Lee, 2013; Penning de Vries, Cucchiari, Bodnar, Strik, & van Hout, 2015), and developing new methods of data analysis for the new

types of data being generated (cf. Amaral, Meurers, & Ziai, 2011). Without doubt, exploration of these issues is essential for contributing to improved second language learning opportunities. However, one area that has received relatively less attention is that of communicating results from external computer-based assessments of second language ability (Jang, 2005; Jang, Dunlop, Park, & van der Boom, 2015). This area is valuable because such assessments, as previously noted, have undergone rigorous development and validation processes, and provide valuable, trustworthy information on second language skills. Communication of this information, as feedback, is important for learners and their learning. Moreover given the opportunity for the fields of language learning and language assessment to grow closer, investigating how learners use feedback is a substantive contribution to both fields. As will be seen below, areas of particular interest are attention, cognition, affect, individual learner background characteristics, and of course usage of feedback.

1.3 Attention and processing in language learning

In order to use feedback about second language skills for second language learning, two key factors are attention to and processing of feedback. Indeed, both these factors are necessary in order to use feedback (Pintrich, 2004), yet little work has been done on exactly what attention and processing looks like in the context of feedback from second language assessments nor their relationships with usage and learners' background characteristics. For example, attention is the focus of an individual's cognition and affect on specific stimuli, and can be measured indirectly through eye tracking (Duchowski, 2007; Just & Carpenter, 1976, 1980; Richardson, Dale, & Spivey, 2006) and other physiological indicators, through self-report, and through tests of recall. Moreover, attention is a prerequisite to processing, yet the type and extent of attention that results in extensive processing can vary depending on a wide variety of factors (Duchowski, 2007; Owen, 2016). Although substantial research exists on how attention varies during reading in general (Richardson, Dale, & Spivey, 2006; Posner, 1980), including second language reading (Kang, 2014; Bisson, van Heuven, Conklin, & Tunney, 2014; Winke, Sydorenko, & Gass, 2013), and written second language assessments (Bax, 2013; Ballard & Lee, 2015; Owen, 2016; Suvorov, 2015), research is not yet available on attention to feedback for second language learning. Given the foundational nature of attention for learners' interactions with feedback, it would seem to be an area demanding some study.

Meanwhile, processing includes (but is not limited to) the conscious and unconscious mental activities that occur within individuals by which novel information is taken in, analyzed, organized, evaluated and integrated into an individual's knowledge, skills and identity. Processing includes cognitive and affective strategies, and is a key component of learning (Pintrich, 2004). External environmental factors such as social expectations, distractions, and consequences can affect processing, as do the affective, cognitive and psychological states and traits of an individual receiving the information.

Within processing, cognitive strategies are the online thinking skills that people use to make meaning of the world around them, and to deal with challenges they encounter. In language learning, cognitive strategies are most often studied in the context of meaning making in the target language, and a factor shown to be key in strategy use is target language proficiency (Lin & Yu, 2015; Phakiti, 2003; Raoofi, Chan, Mukundan, & Rashid, 2014; Zhang, Goh & Kunnan, 2014). However despite the importance of using effective cognitive strategies in self-regulating learning (Pintrich, 2004), there are no studies looking specifically at how learners utilize cognitive strategies while processing assessment feedback on their second language skills – a key point in the self-regulated learning process.

Simultaneously, affective strategies are the online emotional skills that people use to maintain task focus and a secure self-identity when interacting with the world around them. Affect and control of affect have been shown to be closely linked with attention (Wadlinger & Isaacowitz, 2011) and social and academic success (Eisenberg, Sadovsky, & Spinrad, 2005). This feature is particularly relevant in the case of language learning, as there is a close relationship between language and identity (Dörnyei, 2005; Pintrich, 2004) and a known link between anxiety and language learning (Swain, 2013; Yorke, 2011), and the presence of these relationships indicate a high likelihood that affective strategies will play a key role in the mediation of feedback processing on second language skills. However once again, despite the observation that receiving disappointing feedback is a key risk point on the language learning journey (Falout, Elwood, & Hood, 2009; Rose & Harbon, 2013), there is a gap in the literature in terms of the affective strategies that learners use when negotiating feedback in particular.

Finally, the processing outcomes – the affective and cognitive states – that learners take away from their feedback experiences are important for subsequent usage of that feedback (Orsmond,

Merry, & Reiling, 2005; Poulos & Mahony, 2008; Price, Handley, & Millar, 2011). However, there is surprisingly little research available about the nature of and factors impacting critical engagement with and acceptance of feedback, including feedback for language learning, although what research exists indicates that learners' self-regulated learning skills, including targeting of attention and use of cognitive and affective skills, as well as learners' background characteristics, are all important (Furnborough & Truman, 2009; Jang, Dunlop, Park, & van der Boom, 2015).

In summary, feedback holds a key position in the learning process, particularly in self-regulated learning (Hattie, 2008; Hattie & Timperley, 2007), and control and use of attention, cognition and affect is very important in effective self-regulated learning (Pintrich, 2004). Given the lack of existing work bringing these two areas together, in the present study I chose to explore language learners' use of attention, cognition, and affect, both in terms of online processes and subsequent outcomes, when they received computer-based feedback from a second language assessment. The link between these processes and outcomes and subsequent usage of assessment feedback was another key area of interest that would better link the fields of language measurement and assessment-for-learning.

1.4 Learner characteristics and interaction with language learning feedback

Throughout this discussion on the ways in which language learners differ in their usage of feedback, individual characteristics that learners bring to the feedback have been repeatedly raised. In this study, several widely researched learner characteristics were identified as important to include in analyses that explored individual differences in processing and usage of diagnostic feedback. The constructs identified and introduced below were language proficiency, motivation, goal orientation, and beliefs about intelligence.

A key factor in usage of feedback for language learning is the language proficiency of the learner. Cognitive processing is substantially affected by specificity of information, but the effects differ according to skill level, with lower and higher proficiency learners benefitting from different amounts and types of information (Kalyuga, Ayres, Chandler, & Sweller, 2003; McNamara, Kintsch, Songer, & Kintsch, 1996; Shin, Schallert, & Savenye, 1994). This effect has likewise been noted among second and foreign language learners, particularly where feedback is delivered in the target language (Chen & Yen, 2013; Fontanini & Braga Tomitch, 2009;

Segalowitz & Frenkiel-Fishman, 2005). As a result, English language proficiency was an essential construct to include in the study.

Language learning goals and motivations for learning a language are very much a product of an aspirational desire to gain the linguistic and cultural skills – and associated material and social benefits – connected with the language (Dörnyei, 2005). Language learners with stronger desires to gain these skills (and benefits) have been observed to demonstrate greater persistence in learning (Awad, 2014; Kormos, Kiddle, & Csizér, 2011; Matsumoto, 2009; Wildsmith-Cromarty & Conduah, 2015). However there is little work available on how these motivations are associated with usage of feedback on language learning (DePasque & Tricomi, 2015). Moreover, in the present study, participants were adult immigrants to Canada, a group known to be highly motivated to achieve membership in Canadian society despite facing various disadvantages (Derwing & Waugh, 2012; Duguay, 2012; Han, 2009). Therefore, language learning goals and motivations were learner characteristic variables for exploration in the study.

Goal orientation can be defined as the underlying reasons that individuals attempt to succeed during a task (Button, Mathieu, & Zajac, 1996; Elliot & Harackiewicz, 1996; Midgley et al., 1998). In this study, three goal orientations were recognized: mastery goal orientation, in which individuals aim to master knowledge or skills; performance prove goal orientation, in which individuals aim to prove to others that they are capable; and performance avoid goal orientation, in which individuals aim to avoid looking stupid to others. Goal orientation is known to be related to the type of feedback sought by learners (Butler, 1993; Jang, Dunlop, Park, & van der Boom, 2015; Pappachan, 2008; Tuckey, Brewer, & Williamson, 2002; Vandewalle, 2003), and thus was an important set of constructs to include for consideration in the study. Its inclusion was particularly important because little research has been done specifically on the impact of goal orientation on language learning (Gorges, Kandler, & Bohner, 2012; Mantou Lou & Noels, 2016; Nakayama, Heffernan, Matsumoto, & Hiromori, 2012).

‘Beliefs about intelligence’ represents a theory which categorizes beliefs that people hold about intelligence (Dweck, Chiu, & Hong, 1995). Two types of beliefs have been identified: people that believe intelligence is a fixed, innate trait that cannot be changed are said to hold fixed beliefs about intelligence. People that believe intelligence is a changeable, malleable state that can be changed over time are said to hold incremental beliefs about intelligence. Beliefs about

intelligence have been shown to relate primarily with willingness to persist in the face of difficulty (c.f. Yeager & Dweck, 2012) and acceptance of criticism (c.f. Mangels, Butterfield, Lamb, Good, & Dweck, 2006; Zhao, Zhang, & Vance, 2013), both of which are integral aspects of feedback delivery. Moreover, the impact of beliefs about intelligence on language learning has to date received little attention (Mantou Lou & Noels, 2016; Miele, Finn, & Molden, 2011), despite the widely acknowledged existence of language learning aptitude (Carroll, 1993; Dornyei & Skehan, 2003; Robinson, 2013). For all these reasons, it was important that this construct was included in the study.

In summary, individuals' language proficiency, motivation, goal orientation, and beliefs about intelligence are known to impact their learning behaviour. However, coverage is uneven in terms of determining how language learners' characteristics are associated with language learning behaviour and usage of feedback on their language learning. Therefore this topic was identified as a key area to which the study might contribute.

1.5 Purpose and significance of the study

As introduced above, substantial research has shown that learner characteristics – such as motivation, language proficiency, goal orientation and beliefs about intelligence – influence learner actions and future achievement. In addition, affective strategies and cognitive processing strategies have all been shown (see above) to affect how well learners are equipped to use feedback for learning. However, there has been limited attention to the actual affective and cognitive strategies that second language learners use when receiving and processing feedback on their second language proficiency. Similarly, there has been limited attention to the relationships that individual characteristics and processing have with usage of feedback on second language proficiency. Therefore, this study was conducted to address these gaps in the literature.

Moreover, the study offered the opportunity to take a step toward making a cohesive link between assessment design and feedback. To date a misstep appears to exist between the fields of psychometric assessment and assessment-for-learning, with the former directing attention toward the quality of the measure, and the latter directing attention toward the effectiveness of feedback. This disconnect substantially lessens the chances of high-quality information on second language learners' abilities based on language tests being communicated meaningfully

into learning contexts. Given the importance of high-quality information for effective learning (Hattie & Timperly, 2007), it is essential that the two fields broach the divide. It is to this need for considered links between assessment design and feedback that this study also aims to contribute.

As a result of these gaps in the literature and need for bridges between fields, the present study investigated how adult immigrant language learners processed and used feedback based on a cognitively diagnostic model that was delivered by computer. The study also investigated how language learners' processing and usage of feedback differed according to individual learner characteristics. A non-experimental design with mixed methodology was adopted, and multiple data sources were collected, including surveys, eye-tracking data, and stimulated recall and delayed recall interviews. I worked with adult immigrants to Canada who were learning English.

The two main purposes of the study were (a) to investigate how adult immigrant English language learners engage with and process computer-based feedback on their English skills and (b) to explore how these learners use feedback depending on their processing outcomes, psychological characteristics, and English proficiency. These purposes were guided by the following four research themes and associated research questions:

- Theme 1 – Characteristics of adult immigrant English language learners in Canada
 - What are their occupational contexts?
 - What are their English language environments?
 - What are their English language goals?
 - What are typical goal orientation profiles and beliefs about intelligence?
 - What are typical English reading proficiency profiles?
 - How do they perceive their English reading proficiency?
 - What relationships are observable between individual characteristics?
- Theme 2 – Relationship between attention to and processing of feedback
 - To which aspects of feedback reports do language learners report paying attention?
 - To which aspects of feedback reports are language learners observed to pay attention?
 - Which aspects of feedback reports can be recalled by language learners one month later?

- Theme 3 – Language learners’ affective and cognitive interaction with report content
 - What affective and cognitive strategies do language learners report when receiving a feedback report?
 - What processing outcomes do language learners report after receiving a feedback report?
 - In what ways are language learners’ processing outcomes related?
- Theme 4 – Usage of feedback
 - How do language learners use a feedback report to plan their learning?
 - What are the relationships between individual characteristics, attention, processing experience, and usage of feedback?

1.6 Overview of dissertation

This dissertation is divided into six chapters. This introduction is the first chapter, followed by the literature review in the second chapter. The literature review situates the study within existing work, and identifies areas that require further work, to which this study can contribute. The third chapter focuses on methodology, describing in detail the study context and design, participant recruitment and characteristics, data collection activities, instrument development, and analyses. The fourth chapter presents the results of the study under the four themes that organize the research questions. The fifth chapter discusses and synthesizes the results among the four themes and in the context of existing literature. Finally, the sixth chapter summarizes the findings and presents some implications as well as acknowledging the limitations of the study and areas for further research.

1.7 Terminology for the study

Multiple fields of study are brought together in this study, and therefore some terminology may be unfamiliar to some readers, and understood in different ways by others. In Table 1, Table 2, Table 3, and Table 4 below, key terms are named and defined according to how they have been used in this study. While some terms have been introduced in the introduction, all the terms will be discussed in detail in the literature review and methods chapters.

Table 1

Assessment and learning terms

Assessment-for-learning	Assessment-for-learning refers to assessment that is conducted specifically to inform and possibly adjust an ongoing program of learning. In this study, the assessment was conducted for assessment-for-learning purposes.
Self-regulated learning	Self-regulated learning refers to the set of affective, cognitive and metacognitive skills that enables a learner to monitor, evaluate and adapt their learning so as to increase learning efficacy and continue progress.
Feedback and feedback report	Feedback refers to the information attained as a result of some activity. Feedback is necessary to engage in self-regulated learning. In this study, feedback on participants' English reading skills is given in the form of a computer-based report designed to facilitate and scaffold self-regulation of learning. Components of the report included an introduction to the purpose of the test, explanations of skills tested, a comparison of test results and self-assessment results, suggestions to direct learning, and a structured planning opportunity.

Table 2

Language measurement terms

Cognitively diagnostic assessment	The test results reported to learners in this study are based on the results of a cognitively diagnostic assessment. Cognitively diagnostic assessment, also known as cognitive diagnostic assessment, is a multidimensional probabilistic latent class modelling approach that assesses the probability an individual has mastered a pre-identified set of skills. Its purpose is to inform learning by providing diagnostic information on which skills need further work.
Cognitive diagnostic modelling	Cognitive diagnostic models are the statistical models developed for use in cognitively diagnostic assessments. Once a model has been developed, it can be applied to an unlimited number of learners, as long as learners complete the same test on which the model was developed. Cognitive diagnostic models generate probabilities of mastery for each of the skills identified on the assessment, for every test taker. These probabilities are the 'test results'.
(Posterior) probabilities of mastery	Posterior probabilities of mastery are the probability that an individual has mastered a specific skill. In this study they are usually referred to as 'probabilities of mastery' or <i>PPMs</i> . Probabilities of mastery are provided on a 0 to 1 scale and are the 'test results'. In this study, six probabilities of mastery for each of six English reading skills were reported to study participants.

English language proficiency	English language proficiency represents the ability of individuals to use English language knowledge and skills in communicative acts to construct meaning between people. In this study, participants' English language proficiency was estimated from the English reading proficiency test results in two ways –the <i>PPMs</i> and the raw total score of items answered correctly.
Self-assessment	Self-assessment takes place when individuals assesses their own knowledge or skills, often according to pre-specified criteria. In this study, participants self-assessed their English reading skills based on their estimated performance on the English reading proficiency test.

Table 3

Psychological terms

Attention	Attention is the focus of an individual's cognition and affect on specific stimuli. Attention can be measured physically through eye tracking and other physiological indicators, through self-report, and through tests of recall.
Processing	Processing represents the conscious and unconscious mental activities that occur within individuals by which novel information is taken in, analyzed, organized, evaluated and integrated into an individual's knowledge, skills and identity. Processing includes affective strategies and cognitive strategies. Processing can be affected by external environmental factors, as well as the affective, cognitive and psychological states and traits of an individual receiving the information.
Cognitive strategies	Cognitive strategies are thinking skills that people use to make meaning of the world around them, and to deal with challenges they encounter. Cognitive strategies discussed in this study include connecting one's own experiences with new material encountered, critically evaluating new material and experiences, and actively wrestling with and making meaning of difficult material. Cognitive strategies are online thinking skills that are used as situations occur and new information is encountered.
Affective strategies	Affective strategies are emotional skills that people use to maintain task focus and a secure self-identity when interacting with the world around them. Affective strategies discussed in this study include emotional regulation, self-encouragement, downplaying information that would otherwise damage self-esteem, and limiting attention to ego-damaging information. Affective strategies are online emotional skills that are used as situations occur and new information is encountered.
Processing outcomes	Processing outcomes are the affective and cognitive states that people report taking away from an experience. Processing outcomes discussed in this study include sense of overwhelmedness, reflection on one's English skills, trust in the report content, desire to use the report, and feeling in need of help to use the report. Processing outcomes are cognitive and affective states that an individual does or feels after a situation or encounter has ended.

Goal orientation	Goal orientation theory represents underlying reasons that individuals attempt to succeed during a task. People aim to master knowledge or skills (mastery goal orientation), to prove that one is capable (performance prove goal orientation), and to avoid looking stupid (performance avoid goal orientation). Goal orientation can change depending on context, and individuals can hold multiple goal orientations simultaneously.
Beliefs about intelligence	<p>Beliefs about intelligence theory represent beliefs people hold about intelligence. Some people believe that intelligence is a fixed, innate trait that cannot be changed, and are said to hold fixed beliefs about intelligence. Others believe that intelligence is a changeable, malleable state that can be changed over time, and are said to hold incremental beliefs about intelligence. Beliefs about intelligence can change depending on the learning domain and context.</p> <p>In the context of second language learning, beliefs about intelligence theory is frequently manifested through beliefs about language aptitude and its expected impact on potential language learning success. For example, as ‘language aptitude’ is considered a biological characteristic, those who view ‘language aptitude’ as an important factor in language learning may also hold strong fixed beliefs about intelligence in a language learning context. Those who view ‘language aptitude’ as only weakly related to language learning outcomes may hold stronger incremental beliefs about intelligence for language learning.</p>
Intelligence	In this study, ‘intelligence’ refers to the ability of an individual to learn additional languages.

Table 4

Eye-tracking terms

Eye tracking	Eye tracking is a method by which the gaze of an individual is tracked as they look at a screen. Small cameras in the screen are trained to follow the individual’s eyes, and tracking data is linked to screen content. Eye-tracking output can include visual maps of where the eyes were looking, and quantitative data including time and number of visits in specific areas. Note that there is not a one-to-one relationship between an individual’s gaze and an individual’s attention to the viewed material. In this study, eye tracking is used to explore participants’ time spent looking at areas of the report, and to prompt participants’ recollections of what they were thinking when first reading their report.
Traces	Eye-tracking traces are the recordings of the chronological path that an individual’s eyes took across a screen. They are manifested by a line that follows the path of individual’s gaze, punctuated by dots representing fixations (see below).
Fixations	Eye-tracking fixations are the points at which an individual’s eye gaze rests on a specific point. Fixations generally constitute the majority of time an individual spends looking at a screen. Depending on the amount of content on the screen and the detail with which it is viewed, there may be hundreds of fixations within a single screen, and each fixation may be only milliseconds in length.

Chapter 2: Literature review

2.1 Introduction

This literature review introduces and discusses several key concepts utilized in this study. The place of feedback in language learning is reviewed, and cognitively diagnostic assessment and its uses for language learning are introduced. The need for validation processes for language assessment feedback is discussed. Key factors related to feedback usage are introduced and reviewed in the context of language learning. Factors reviewed are self-regulated learning skills, including cognitive and affective strategies, and individual learner characteristics, including second language learning environment, goal orientation and beliefs about intelligence, second language proficiency, and self-assessment skills. Throughout the review, gaps in the literature to which this study will contribute are highlighted.

2.2 The place of feedback in language learning

Educators often seek to utilize assessment practices such that students benefit in terms of learning and personal development. Broadfoot et al. (1999) argued for the development of ‘assessment-for-learning’, which they defined as “the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there” (pp. 2-3). The benefits of such assessment practices have been shown to include substantial increases in achievement (Black & Wiliam, 1998; Parr & Timperley, 2010) and self-regulated learning and self-efficacy (Clark, 2012).

As Broadfoot et al. (1999) furthered in their argument, absolutely crucial to effective implementation of assessment-for-learning practices, including second language learning, is the creation, communication and use of evidence that informs learners and their teachers “where the learners are in their learning, where they need to go and how best to get there” (pp. 2-3). This evidence is generally referred to in the literature as ‘feedback’ (Wiliam, 2011). The literature on educational feedback is vast (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008), and Bangert-Drowns et al. (1991) avoid even defining ‘feedback’ as a single term, noting that feedback “may not be a unitary phenomenon” (p. 214). Instead, they discuss how feedback can differ by intentionality, mode of delivery, target (purpose), and content.

Bangert-Drowns et al.'s statement about the broad possibilities of feedback remains relevant, evidenced in the wide range of operationalizations of 'feedback' in the literature. Like Bangert-Drowns et al., many studies use feedback mode, purpose and content as a de facto definition of feedback rather than defining the term itself. In this sense, feedback might be verbal, written or observed (Biber, Nekrasova, & Horn, 2011), might originate from teachers, peers, computers or instructional materials, might be immediate or delayed (Clariana, Wagner, & Murphy, 2000), and might provide knowledge of result, knowledge of correct response, elaborated feedback (Dihoff, Brosvic, Epstein, & Cook, 2004; van der Kleij, Eggen, Timmers, & Veldkamp, 2012) or worked examples (Kopp, Stark, & Fischer, 2008). Feedback might be corrective and evaluative or descriptive and process-oriented (Montgomery & Baker, 2007), may focus on the learner or the work, and may be affectively positive, negative or neutral (Carless, 2006; McLaren, 2012). Communicative aims of feedback may be to motivate, evaluate, criticize or point forward (Lee, Lim, & Grabowski, 2010).

However in language learning, research on feedback has focussed on the potential learning benefits of immediate versus delayed feedback, and the relative strengths of knowledge of result, knowledge of correct response, and elaborated feedback, including how language learners interpret written feedback. These three topics are now reviewed.

In terms of immediate versus delayed feedback, a common focus of research on feedback for second language learning, timing has not been shown to consistently affect language acquisition. Opitz, Ferdinand and Mecklinger (2011) found that when learning artificial grammar rules, learners acquired the rules faster with immediate feedback. They argued that immediate reinforcement is necessary for language learners to acquire grammar inductively, otherwise they simply have no way of observing the rules of the grammar. However, Nakata (2015) found that timing of feedback had little effect for English learners in Japan when they learnt a set of English words, and Fu and Nassaji (2016) found that learners adopted corrections from their teacher more often when it was delayed by a few turns.

Meanwhile, Nassaji (2016) conducted a literature review of research in interactional feedback, which is when language learners receive natural forms of correction during interaction with others, and found the benefits of interactional feedback itself are positive, although impact depends on quality of delivery and learners' ability to notice correction (cf. Mackey, 2006).

Indeed, Ellis, Loewen, and Erlam (2006) noted that for English language learners who participated in their study, explicit feedback in the form of metalinguistic explanations was more effective than implicit feedback (recasts) for learning a piece of English grammar.

In terms of knowledge of results, knowledge of correct response, and elaborated feedback, language learning research has focused on acquisition of language knowledge such as specific grammatical rules and vocabulary. Discussion of the benefits of corrective feedback has been particularly extensive in research on second language writing development (Bitchener & Ferris, 2012; Ferris, 2010; Murphy & Roca de Larios, 2010; Truscott & Hsu, 2008). In general, elaborated feedback of various types is associated with the greatest uptake of accurate written language. For example, Sheen (2007) noted that metalinguistic explanations (a type of elaborated feedback) were associated with greater increase in written accuracy than correction only. Likewise, Van Beuningen, de Jong, and Kuiken (2012) observed that direct corrections of errors (a form of knowledge of correct response feedback) were more effective than knowing what type of errors had been made (a type of knowledge of result feedback).

However, research outside the field of language learning indicates that learners often interpret written feedback on assignments (including both error correction and extended comments) in ways differently than intended by feedback providers. For example, Carless (2006) and Rae and Cochrane (2008) both found that university students often only accurately understood the intended messages in feedback through clarification with the teacher. Lea and Street (2000) additionally found that university tutor and student interpretations of the same comments were often different. One of the greatest difficulties in interpreting feedback appears to be lack of detail that would facilitate understanding (Cramp, 2011; Rae & Cochrane, 2008; Walker, 2009; Weaver, 2006).

In addition, it is important to note that students emerge in research on feedback as active participants in negotiating feedback, including for language learning. Orsmond, Merry, and Reiling (2005) report that students actively use feedback to enhance motivation and learning, encourage reflection and clarify understanding. Moreover, Fernandez-Toro and Hurd (2014) developed a detailed model exploring how language learners learning via distance study engage with and use feedback. They propose that while usage differs according to the characteristics of the learner, learners use feedback as a source of information on the progressive realisation of

their desired identity, as well as direct feedback on their language skills. This perspective of feedback as source of motivation and reflection stands in contrast to an understanding of feedback in which students simply aim to improve learning, and indicates that when students struggle to understand feedback (cf. Carless, 2006; Chanock, 2000; Higgins, Hartley, & Skelton, 2001) substantial benefits are being lost in addition to limiting skills acquisition.

Finally, while second language error correction for grammar and vocabulary, either in isolated study or within the productive skills of speaking and writing, has received substantial attention (Ellis, Loewen, & Erlam, 2006; Li, 2010; Mangels, Butterfield, Lamb, Good, & Dweck, 2006; Spada & Tomita, 2010), much less attention has been paid to feedback on higher level second language skills such as the use and control of discourse structure and turn taking, communicative success, inferencing and comprehension. Particularly lacking is work on feedback on listening and reading skills, although this issue has been noted (Field, 2008; Harding, Alderson, & Brunfaut, 2015).

Moreover, a large literature in diagnosing first language lower and higher order reading skills such as phonological awareness, letter and word recognition, summarizing, inferencing and evaluating already exists to draw on (Bartlett, Dowd, & Jonason, 2015; Snow, Burns, & Griffin, 1998), although the focus of attention has been on identifying important skills rather than on how to deliver feedback to learners, perhaps because it is generally assumed that teachers and parents will use assessment results rather than the typically young learners. For example, Lonigan, Burgess and Anthony (2000) conducted a longitudinal study to identify emergent reading skills that were associated with subsequent reading skills. The National Reading Panel (2000) was a widely praised and reviled review of the components of learning to reading and associated effective teaching practices. Neuman and Dickinson (2001) and Dickinson and Neuman (2006) are comprehensive reviews of current knowledge on emergent and early reading skills, including their component subskills and the sociocultural factors that nurture their growth. Finally, Bartlett, Dowd, and Jonason (2015), Dubeck and Gove (2015), UNESCO-UIS (2016) and others are currently expanding the discussion of what early reading skills to measure, and how to best measure them, to low-material resource contexts in economically developing countries.

Regarding why reading and listening have received less attention in work on feedback for learning, Mendelsohn (2006) notes that foreign language listening is usually tested repeatedly

rather than actually taught as a process. The same might be said, to a lesser extent, of foreign language reading (where the language learner is already literate), although how to teach reading to young second language learners (who are also learning to read) has long been a high priority research area in some countries (Melby-Lervag & Lervag, 2014, Snow, Burns, & Griffin, 1998). Widespread use of ‘testing’ as ‘teaching’ might be one reason for lack of work on feedback delivery, as in these situations getting the answers correct/incorrect on a test is seen as ‘knowledge of correct response’ feedback. A related challenge is, due to the internal nature of listening and reading, deriving useful information about reading and listening skills that can be used in feedback is necessarily inferential and therefore challenging. However, work in diagnostic language assessment is beginning to explore measurement methods for generating and providing feedback on second language reading and listening skills (Harding, Alderson, & Brunfaut, 2015), to be discussed further in the next section.

One final observation is that research on feedback for language learning, as shown above, has to date generally viewed receiving feedback as a manipulable variable with conditions to evaluate rather than a process influenced by multiple psychological and social factors. It is likely that without accounting for the conditions under which, and the processes by which, feedback is picked up and processed, the evidence of impact for feedback on language learning is likely to continue to be equivocal. Current exceptions are the work of Jang (2005), Fernandez-Toro and Hurd (2014), Jang, Dunlop, Park and van der Boom (2015), and Wagner (2015). Jang (2005) provided diagnostic feedback on reading skills to English language learners, and found learners valued the feedback, but that effective usage was subject to multiple contextual and individual factors. Fernandez-Toro and Hurd (2014) specifically sought to understand the cognitive and emotional process by which learners engaged with feedback on their language learning, and proposed a model based on their observations. Jang, Dunlop, Park and van der Boom (2015) shared feedback with 11- and 12-year old students concerning their reading skills. They found that psychological characteristics of the learners, including goal orientation, beliefs about intelligence and self-efficacy, as well as reading proficiency and parental attitudes to learning, all affected how children dealt with their feedback. Wagner (2015) made similar findings with high school immigrant English language learners, and additionally noted the significant influence of the use of feedback by students’ teachers.

Clearly, feedback is an integral part of second language learning, both in structured classroom contexts and in terms of learners' ability to monitor and modify their own language. In fact, it can be argued that due to the social nature of language use, feedback is particularly important for developing the ability to use a second language appropriately. However, despite comprehensive attention to certain types of feedback and their impact on improving second language proficiency, little attention has been paid to exactly how language learners process feedback, and how individual characteristics impact processing and usage of feedback. Work indicating how these issues might be effectively addressed will be surveyed further later in the review. However, attention now turns to an assessment approach known as cognitively diagnostic assessment, and its applications in language learning, as this approach is increasingly being utilized as a means to address the aforementioned challenges of generating language assessment results that offer more than 'knowledge of correct response' or a general estimate of proficiency.

2.3 Uses of cognitively diagnostic assessment for language learning

Cognitively diagnostic assessment (CDA: Rupp, Templin, & Henson, 2010; Tatsuoka, 2009), the assessment approach used in the psychometric model from which this study's feedback is created, is one way that the need for substantive yet efficient information on language skills can be met. CDA brings together cognitive psychology and educational measurement to provide diagnoses on skills of interest (Leighton & Gierl, 2007). First, using theories of language acquisition, cognitive skills required to respond to items correctly on the test are identified, for example using vocabulary and syntax knowledge, extracting explicit information, connecting and synthesizing, and making inferences (Li, 2011). Diagnoses on these skills are then obtained by using cognitively diagnostic modelling (CDM) procedures to provide estimates of mastery on these pre-identified cognitive skills. Based on these mastery estimates, feedback systems can be constructed that present descriptive diagnostic information tailored toward the learning needs of an individual. In doing so, CDA provides the opportunity for psychometric assessment methods to contribute to assessment-for-learning practices by offering information that is useful for adjusting instruction and learning goals.

One major aspect in which CDA differs from psychometric approaches such as classical testing theory and unidimensional item response theory is the assumption that the construct measured by the cognitively diagnostic model is multidimensional (DiBello, Stout, & Roussos, 1995). That is

to say, in CDA the assessment purposely aims to assess multiple skills simultaneously, rather than assessing a single skill such as general math or language proficiency. CDMs enable this by allowing skills to be correlated, thereby allowing for representation of multidimensional constructs.

Moreover, whereas popular psychometric approaches to scaling ability, such as Rasch modelling, approach ability as a latent trait and place individuals on a single ability scale (DeMars, 2010), or in the case of multidimensional item response theory, as multiple latent traits with multiple ability scales (Yan, Almond, & Mislevy, 2004), a CDM is a probabilistic latent class modelling approach. These types of model predict the chances that an individual has mastered a selection of skills as defined within the cognitive model.

The strength of using a CDM is that specific strengths and weakness in required skills for a topic can be identified in one assessment, rather than potentially requiring several assessments that each assess a different dimension. Methodologically, CDMs are similar to multidimensional latent trait modelling methods such as multidimensional item response theory because they produce information on multiple skills, but CDMs produce probabilities of skill mastery, whereas multidimensional item response theory methods produce scaled measures of proficiency on each skill. The ability of both methods to deliver fine-grained descriptive information to individuals is believed to promote a conceptual shift toward assessment-for-learning, and presents a potentially invaluable opportunity for empirical assessment methods to contribute substantively to the movement toward integrated assessment and learning practices.

CDMs are a confirmatory modelling approach based on a Q-matrix (Rupp & Templin, 2008). The Q-matrix is a table of skills by items, which records the skills required for success on each item in the assessment, and is the method by which inter-skill relationships are defined. Therefore, the theoretically sound development of the Q-matrix is critical to ensuring the assessment reflects individual abilities accurately, and requires a combination of expert judgement, application of relevant content-based research and statistical verification (Buck & Tatsuoka, 1998; de la Torre, 2008; Jang, 2005; Kim, 2015; Sawaki, Kim, & Gentile, 2009).

Once the Q-matrix is developed, parameter estimates are produced using a sample of the relevant population, and relevant software. The sample must be large; about 500 individuals is an expected minimum (Jang, personal communication, October 2014). Modelling choices must also

be made when specifying the model. Whether a compensatory approach or non-compensatory approach should be adopted is one of the major choices; a compensatory approach assumes that only some of the identified skills need to be mastered to correctly answer the item, whereas a non-compensatory approach requires all the specified skills to be mastered in order to correctly answer the item. Other significant choices regard decisions about how slippage (when high achieving test takers get an easy item wrong) and guessing (when low achieving test takers get a difficult item right) will be dealt with.

After the CDM is developed, including making any necessary revisions to the Q-matrix, the resulting item parameters can be applied to any individual from the relevant population to estimate the likelihood of skills mastery. From these estimations of skill mastery, feedback can then be developed that will inform learning and instruction.

The field of second and foreign language assessment stands to benefit substantially from developments in CDA, as a large majority of language practices are arguably multidimensional. Moreover, as noted earlier, there has been limited attention to developing diagnostic information about second language reading and listening skills due to the receptive nature of the skills, but CDA is designed to identify multiple latent skills. Therefore unsurprisingly, the language assessment field is demonstrating considerable interest in CDA and CDM development.

Due to difficulties obtaining necessary sample sizes for assessments integrated into classroom teaching and learning, to date work in cognitively diagnostic language assessment has focussed on developing CDMs using large-scale assessments (Buck, Tatsuoka, & Kostin, 1997; Buck & Tatsuoka, 1998; Chen & Chen, 2016; Chen, Gorin, Thompson, & Tatsuoka, 2008; Jang, 2005; Kim, 2015; Lee & Sawaki, 2009; Li, 2011; Li, Hunter, & Lei, 2016; von Davier, 2008). For example, Li (2011) developed a CDM using the Fusion Model (also used in this study's CDM) for the Michigan English Language Assessment Battery (MELAB) reading test, while Li, Hunter and Lei (2016) evaluated the performance of several CDMs on the MELAB reading test. Lee and Sawaki (2009) compared CDMs using reading and listening papers from two versions of the TOEFL iBT® test. Kim (2015) also developed a CDM based on the reading paper of a placement test used for an adult ESL program, which was not a large-scale assessment, although it was still reverse-engineered.

Despite this increasing body of work, Alderson (2010) and Jang (2009, 2010) both note that large scale second language proficiency tests that were originally developed for ranking individuals on an achievement scale generate a less-than-ideal amount of information to diagnose skill mastery. However, Chen and Chen (2016) concluded that (first language) reading tests not designed for CDA purposes could be retrofitted to a CDM in a statistically justifiable manner. More crucially, all the studies above developed CDMs and Q-matrices, but they did not investigate how feedback delivered from these models was received by language learners. In the field of second language learning, the only available work looking at how learners use CDM-derived feedback is still Jang's (2005) study which considered how English language learners used feedback on second language reading proficiency based on a CDM. In the study, Jang found that the diagnostic information offered by such feedback was potentially very useful to language learners and their teachers. However, Jang noted that "the potential of skills diagnosis is subject to the extent to which a set of skills can be sufficiently specified and further to the degree to which it is integrated into curriculum" (p. iv). Thus, questions remain about the ways in which learners' interaction with and usage of feedback varies according to the background characteristics learners bring to the feedback, as well as the environments in which they are learning.

2.4 Language learning feedback validity

As is made clear above, comprehensible, relevant feedback on second language skills is a crucial part of second language learning. Moreover, the value of a cognitively diagnostic assessment approach lies in its ability to inform subsequent learning and instruction by providing meaningful, substantive information on learners' relative strengths and weaknesses in the language skills tested. Regardless of the assessment method, any feedback provided must be presented at a level of detail and in a manner that is meaningful and useful to learners so that future learning is supported (Hattie & Timperley, 2007). As a result, it may be argued the quality of the feedback for any assessment intended to assist language learning is integral to the validity claims of the assessment itself. Within this argument, if the feedback does not facilitate application of assessment findings to future learning, the entire purpose of the assessment comes into question. However, validation and test design approaches developed for second language assessment have not given feedback issues substantive attention and moreover, multiple factors also appear to influence the effectiveness of feedback (Jang, 2005; Hurd & Fernandez-Toro, 2014), indicating

that the validity claims of feedback may be complex. Thus, issues of test validity arguments are now addressed.

In educational assessment, validity traditionally refers to the interpretation of test scores in order to make inferences about abilities in a manner justifiable and appropriate to the information available, the context of the test taker, and the purpose of the test (Kane, 1992). Importantly, the notion of validity is a fluid concept that very much depends on a particular assessment and a particular context (Cumming & Berwick, 1996; Huff & Sireci, 2001). That is, an argument for assessment validity is based on the extent to which its design meets the needs of a specific context and group of test takers. For example, a test of English listening proficiency for elementary school learners studying English as a medium of instruction, despite being a ‘valid’ test in this context, might not be valid for use by adults applying for work visas in English speaking countries. Understanding validity as a dynamic, shifting process frees assessment developers and users to adopt an approach to validation that emphasizes responsible development and ethical usage of second language assessments on a fit-to-use basis. Therefore it is often more accurate to discuss validity in terms of validation processes rather than as a static term.

Widely cited work on validation published since Messick’s (1989) seminal article asserting unity of validation processes, in which Messick argued that various previously defined types of validity were all complementary and related aspects of a unitary idea ‘validity’, supports this view of validity as a process. These works variously prefer to talk in terms of building an argument (Kane, 1992), gathering evidence (Mislevy & Haertel, 2006; Weir, 2005) or supporting a case (Bachman, 2005) for using an assessment. Kane (1992) developed an approach which viewed validation processes as building arguments to support the interpretation of test scores in the way planned, based on Toulmin’s (1958, 2003) argument structure to this approach consisting of the elements of data, claim, warrant (reason claim is justifiable), backing (evidence backing warrant), and rebuttal (any situations in which a warrant becomes unjustified). Mislevy and Haertel (2006) framed validation processes in terms of ‘evidence-centred design’, and incorporated Kane’s work into a different structured process in which evidence is gathered based on examples of performance on a skill, from which generalizable inferences about an individual’s skills are derived, and arguments developed to support the inferences. Weir (2005)

also viewed the validation process as evidence gathering, but he adopted a socio-cognitive approach in which language use is viewed as a social medium, test taker characteristics are accounted for, and context validity and consequential validity take an equal status with the more cognitively oriented theory-based validity, scoring validity, and criterion-related validity. Finally, Bachman (2005) built stronger links between score inferences and test use/consequences, arguing that building ‘a convincing case that the decisions we make are defensible and supporting that case with credible evidence are the two components of the validation process.’

This perspective of validation as an ongoing process that constitutes principled gathering of evidence to justify an assessment’s design and use in a particular context is also applicable when creating a validity framework for feedback. However existing validity and test design frameworks make little to no mention of the role of actual feedback in assessment, although they do address the wider issue of test use. For example, Mislevy and colleague’s evidence-centered design framework is a clear example of validity principles applied to test development. However score reporting is mentioned but not elaborated on, and feedback receives no explicit reference. Similarly, Fulcher and Davidson’s (2007) effect-driven design approach to language assessment, which argues that assessment development should aim to maximize assessment effects, lays out key concepts relevant to feedback validation processes but specifies details only for testing. For example, in defining effect-driven testing Fulcher and Davidson offer limited guidance for direct application of the approach to designing feedback systems.

As general second language assessment validity and test design frameworks relegate feedback to delivery of (validated) results, no frameworks are directly available for use in considering the validity of an assessment’s feedback, and how it complements the validity of the assessment as a whole. Such a framework would be valuable for developing feedback appropriate for the learning purposes of a specific assessment, and for evaluating the strengths and weaknesses of feedback delivered depending on the assessment and instructional contexts, and learner characteristics.

Indeed, even within the educational feedback literature, a fully developed framework for assessment-for-learning feedback validity in general has yet to be developed. However valuable work exists that outlines general principles of educational feedback (c.f. Nicol & Macfarlane-Dick, 2006; Rae & Cochrane, 2008; Shute, 2008), and various frameworks describe how

feedback can be more or less effective (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Clark, 2012; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Narciss, 2004; Shute, 2008, Yorke, 2003) or describe feedback options (Kluger & DeNisi, 1996; Mason & Bruning, 2001; Shute 2008). An important contribution was recently made by Fernandez-Toro and Hurd (2014), who developed a model of how independent adult foreign language learners engage cognitively and emotionally with feedback. By integrating this work with educational assessment validation approaches, a validity framework for assessment-for-learning feedback is proposed.

In terms of design framework, this study adapts Fulcher and Davidson's (2007) effect-driven assessment design principles to feedback design. An effect-driven design approach posits that key to establishing a validity argument is that the assessment (or in this case, feedback) should maximize its intended effects. Arguments must be constructed to justify that the assessment (or in this case, feedback) achieves these effects, and each argument consists of defended claims. For this study, the claims are identified by reference to what contextual conditions and feedback content the literature reports is required to achieve the goals of feedback. Evidence must then be gathered to defend these claims and establish a validity argument for the assessment-for-learning feedback system in question.

As an effect-driven approach, the intended effects of the present study's feedback should be reiterated. These effects are to motivate and assist continued English language learning, specifically concerning reading skills, in an individual. For this study's context, feedback should be descriptive, cognitively-based information on current English reading proficiency and specify recommended next steps for further learning. In addition, Nicol and Macfarlane-Dick (2006) identified through literature review several observable outcomes that have been shown to facilitate learning, which might be considered 'intermediate' outcomes on the road to learning. These identified outcomes were student self-assessment (reflection) about learning, teacher and peer dialogue around learning, and positive motivation and self-efficacy within learners. Students should also demonstrate an understanding of how to make progress on their learning (Birenbaum et al., 2006; Clark, 2012; Hattie & Timperley, 2007). These observable outcomes are not the ultimate intended effect of learning progress, but they provide evidence that the final desired outcome is likely to occur. With these desired effects in mind, I will now consider what evidence exists regarding expected cognitive processes and processing outcomes after language

learners receive feedback that generates the intended effects of motivation and assistance in continued English language learning, specifically concerning reading skills.

2.5 Self-regulated learning in language learning and its relationship with feedback

The role of learners' behavior in influencing the learning process is recognized as both substantial and key to successful learning (Hattie & Timperley, 2007). The ways in which learner behavior influences learning are also recognized to be many and related in complex ways (Pintrich, 2004). The term loosely used to identify this set of behaviors is self-regulated learning, but there is no one empirically confirmed model that defines the aspects of learner behavior and their interrelationships (Puustinen & Pulkkinen, 2001) nor is there such a model for language learning specifically. However, consistent features in models of self-regulated learning are an iterative cycle of preparation, doing, and reflection (Dörnyei, 2005; Pintrich, 2004; Puustinen & Pulkkinen, 2001; Zimmerman, 2002), and competencies covered can include goal setting, planning, monitoring, reflecting, help seeking, compensation strategies and application of learning strategies.

In the present study Pintrich's four-phase framework of self-regulated learning (Pintrich, 2004) was adopted because of its use of social cognitive theory (Bandura, 1989) and subsequent integration of affective, motivational and contextual factors such as goal orientation and beliefs about intelligence. Pintrich's four phases are first forethought, planning and activation, then monitoring, then control, and finally reaction and reflection. As with all theories of self-regulated learning, these activities are iterative, not necessarily linear, and constantly used during learning. Across the four phases, Pintrich identified four areas for regulation: cognition, motivation/affect, behavior, and context. Within each of the resulting 16 cells, Pintrich then described the individual actions that take place to help or hinder learning. For example, the first phase of cognition involves goal setting and activation of prior content and metacognitive knowledge. Meanwhile the first phase of motivation and affect involves goal orientation adoption, judgements about self-efficacy, establishing perceptions of task difficulty, assigning value to the task, and activating interest. The second phase of both cognition and motivation and affect involves online metacognitive awareness and self-monitoring of cognition/motivation and affect. The third phase for both domains then involves selecting and adapting appropriate strategies. Finally, the fourth phase for cognition involves making judgments and conclusions, while the

fourth phase for motivation and affect involves emotional reactions and attributing these reactions to a source.

However, Pintrich (2004) himself acknowledged that empirical research has shown that observing all the separate constructs is difficult given that many processes are automatic and occur almost simultaneously. He argued that aspects of self-regulated learning will be observable depending on the task, and it is therefore perfectly acceptable to observe self-regulated learning at a larger grain size than that specified in his model, based on the nature of the task.

The present study looked at how language learners used feedback reports from a language proficiency assessment. Therefore, in terms of grain size, the study focused on how language learners used the report, how existing language skills, motivation and affect were associated with this usage, and how this varying usage was observable through differing usage of cognitive and affective strategies to deal with and use the feedback on their language learning.

2.5.1 Cognitive strategies while dealing with second language material

Cognitive skills are an essential construct to consider when investigating educational feedback usage because the successful processing and application of the feedback requires application of cognitive skills (Pintrich, 2004). Unfortunately, regarding types of cognitive strategies that language learners employ while processing feedback on their target language, no prior research on this topic was found. However, there is substantial work on the use of cognitive and metacognitive strategies in general language learning practice and usage as well as in second language reading comprehension. This latter area of work is relevant because the feedback in this study is delivered as figures and written text, so work indicating how strong second language readers use cognitive strategies is relevant to this study, and is discussed below.

Regarding use of cognitive strategies in second language learning in general, it is worth noting that more proficient language users consistently report more use of cognitive strategies for all skills (Phakiti, 2003; Raoofi, Chan, Mukundan, & Rashid, 2014; Sun, 2013; Zhang, Gu, & Hu, 2008). Although this finding is about language learning in general and not specific to dealing with feedback, it remains relevant to the study in that receiving feedback on language skills is part of second language learning. These studies tend to assume that more proficient language users are strong language learners, and therefore strong language learners use more cognitive strategies. However, given that many low proficiency language users in all contexts and of all

ages eventually become high proficiency language users, the findings rather suggest that there are issues related to working memory and cognitive load impeding lower proficiency language users' uses of metacognition. Zhang, Goh and Kunnan (2014) do however argue through use of structural equation modelling that metacognitive and cognitive strategies appear to have limited effect on higher test performance of English as a foreign language readers.

Regarding second language reading in particular, findings reflect the wider field, and are relevant to this study because the test was of second language reading, and the report was given in English – the target second language. Saengpakdeejit and Intaraprasert (2014) interviewed Thai learners of English and noted a wide variety of reading strategies, with more proficient readers reporting more frequent use of strategies. Lee-Thompson (2008) found that American learners of Chinese use top-down and bottom-up reading processes when reading, as do all readers, with an orthography-specific area of particular challenge being identifying word boundaries. Lau and Chan (2007) found that strategy training with Hong Kong young learners struggling to learn to read Chinese was effective at improving and maintaining reading achievement, although the improvements were not observable in other subjects.

To investigate further the question of whether higher proficiency equates to being a strong language learner, Lin and Yu (2015) conducted think aloud interviews with Taiwanese learners of English while reading Chinese and English. They found similar usage of metacognitive strategies in both languages, with wider variation in strategy usage for English than Chinese. They also found that less proficient English readers used language-oriented strategies such as translation, while more proficient English readers used more higher-order cognitive strategies such as paraphrasing, using contextual clues, and asking themselves questions. Their conclusion was that language proficiency does indeed have an effect on the cognitive and metacognitive skills available for use to the learner. This finding indicates that language proficiency is an important factor to consider when observing cognitive processing of feedback on second language skills.

2.5.2 Affective strategies while dealing with language learning challenges

In addition to the centrality of cognition in self-regulation of learning, affect is a second key area for control. It is believed that the ability to regulate affect using a variety of strategies facilitates the selection of appropriate actions and choices, and reduces oversensitivity to negative or

positive stimuli that will result in poor decisions (Wadlinger & Isaacowitz, 2011). Note that the ability to regulate affect is strongly related to social success and somewhat related to language skills and academic success – but as language skills, social ability and academic success are related, there are also recognized mediation effects (Eisenberg, Sadovsky, & Spinrad, 2005; Monopoli & Kingston, 2012). As language, academic and social success are all very much intertwined in second language learning, particularly for immigrants, these mediation effects are potentially particularly strong for immigrant second language learners.

In terms of how affect and affective strategies interact with the psychology of individuals, Wadlinger and Isaacowitz (2011) conducted a literature review about the relationship between regulating attention and regulating emotion. Attention is shown to have strong links with affect, and individuals moderate their attention as part of effective emotional regulation. They also note that attention control can be improved, with ensuing improvement in emotional regulation. Huang (2011) conducted a meta-analysis that noted a positive relationship between mastery goal orientation and positive affect, and between performance avoid goal orientation and negative affect. No linear relationship was observable between affect and performance prove orientation. Finally, Ertac's (2011) work in economics noted that unless individuals were highly confident about their skills, receiving negative feedback on their skills led to a greater decrease in positive self-image than was warranted.

In the case of language learning, the task is not simply to acquire new skills, but also to expand one's self-identity to incorporate the new ways of thinking implicit in acquiring new language, and to develop some sort of membership in the language cultural group (Dörnyei, 2005; Swain, 2013). As language and identity are so closely linked, emotions in language learning can be particularly strong. Add the fact that language learning requires ongoing motivation and effort, both of which can be negatively impacted by negative affect (Ertac, 2011; Rose & Harnon, 2013), and the potential for emotional regulation to impact language learning is high.

However, the role of affect in language learning is not straightforward. For example, Rose and Harbon (2013) found that foreign language learners of Japanese kanji had most trouble regulating their emotions and maintaining motivation at higher levels of proficiency, which the authors attribute to learners seeing limited progress. In contrast, Falout, Elwood and Hood (2009) found that Japanese learners of English with lower proficiency were least likely to control their

affect and maintain motivation. Meanwhile, Pilipovic and Glusac (2016) observed that use of affective strategies in language learning was highest among primary school age children, and then dropped off.

Moreover, although the second and foreign language learning fields have investigated anxiety in quite some detail (Swain, 2013; Yorke, 2011), and affect is recognised as a significant part of self-regulated learning (Dörnyei, 2005; Pintrich, 2004), there is little research that explores how learners' emotions modulate during language learning, and specifically how learners deal with feedback from an affective perspective. An exception is a study that followed a beginner learner of Portuguese through the first eight weeks of study (Garrett & Young, 2009). The study notes the range of emotions reported and that the learner negotiates and regulates emotion alongside ongoing motivation and during various cognitive experiences. In addition, Mantou Lou and Noels (2016) found that emotional responses to language learning feedback were related to goal orientation and beliefs about intelligence. They found that fear of failure and more helpless responses were related to performance orientations and fixed beliefs about intelligence.

2.5.3 Interaction of cognition and affect

An issue that involves the interaction of cognition and affect during feedback processing is that of user perception of the feedback. The extent to which users perceive educational feedback as useful, comprehensible and trustworthy has been shown to impact the extent to which feedback is utilized, regardless of the objective quality of information contained in the feedback (cf. Carless, 2006; Corbalan, Paas, & Cuypers, 2010; Lee, Lim, & Grabowski, 2010; McLaren, 2012; Pokorny & Pickford, 2010; Timmers & Veldkamp, 2011). In particular, positive perceptions of feedback have been shown to relate to increased learner motivation (McLaren, 2012; Price, Handley, & Millar, 2011), while negative perceptions are related to ignoring feedback (Poulos & Mahony, 2008; Rae & Cochrane, 2008). This body of evidence indicates that user attitudes toward feedback should be positive to facilitate processing and implementation of feedback.

Fortunately, students are generally positive about educational feedback received (Miller, 2009; Pitts, 2005) and place especially high value on personalized feedback (Orsmond, Merry, & Reiling, 2005; Walker, 2009). However, perceptions about feedback appear influenced by multiple contextual factors (Poulos & Mahony, 2008; Weaver, 2006) such as attitudes toward the instructor providing the feedback, the way in which feedback is communicated, how feedback is

integrated into learning (McLaren, 2012) and student learning orientation (Dweck, 1986; VandeWalle, 2003).

Moreover, the above body of research investigates learner acceptance and engagement as an (assumed positive) outcome, and does not address the cognitive and affective processes by which acceptance and engagement are achieved (or not). Indeed, despite the importance of metacognitive control and use of reflection in models of self-regulated learning (Dörnyei, 2005; Pintrich, 2004) there is surprisingly little research available about the nature of and factors impacting critical engagement with and acceptance of feedback, including feedback for language learning. Jang, Dunlop, Park, and van der Boom (2015) found that young learners holding dominant mastery goal orientations were more likely than those holding dominant performance orientations to critically engage with feedback on their reading skills rather than simply accept it. Additionally, Furnborough and Truman (2009) found that adult beginner foreign language learners of European languages who compared feedback with their own experience using the language, and integrated feedback into their learning, were also the most confident learners and able to maintain motivation over time.

2.5.4 Monitoring cognition of feedback on second language learning

It is essential that learners process the information provided in feedback, and in order to process it, the learner needs to pay attention, although not necessarily consciously (Duchowski, 2007). Where learners fail to process information, even where the information is good quality (see above), feedback cannot be integrated into the learner's knowledge frameworks, and therefore cannot be retained and applied to progress learning. Two ways to access learners' attention is by tracking their gaze, and by asking them to report it.

Just and Carpenter (1976, 1980) were early leaders in the relationship between cognitive processing, attention, and visual gaze. They noted that longer fixations occurred when there were greater processing loads, and that differences in working memory might result in differing abilities to process cognitive load. In time, fixations were identified as one of the key physical measures of gaze (Duchowski, 2007). Research in the 1980s also yielded that while attention and visual gaze are highly related, attention can focus on areas outside the direct gaze, or outside the gaze altogether (Richardson, Dale, & Spivey, 2006; Posner, 1980; Rayner, 1998).

As a result, although the precise physical mechanisms of visual gaze remain under some debate (Duchowski, 2007), eye tracking has become an established method for exploring attention. The first work was done with reading in a first language (Just & Carpenter, 1976, 1980), but the field of second language learning has taken up use of the emerging technology (Kang, 2014; Bisson, van Heuven, Conklin, & Tunney, 2014; Winke, Sydorenko, & Gass, 2013). For example, Kang (2014) investigated reading strategies among first and second language readers and found that the second language readers read much more slowly, but the two groups were otherwise alike in attention distribution and reading comprehension, indicating that both groups were utilizing the same cognitive processes.

Recently, language testing researchers have begun to explore via eye tracking how language test takers engage with test tasks. Bax (2013) found that eye tracking yielded valuable information about how stronger and weaker adult test takers interacted with some second language reading test items, while Ballard and Lee (2015) made the same findings with young learners. However, based on research into stimulated recall interviews and cognition with adults, Owen (2016) suggested that care should be taken to elicit recall from research participants before showing them their fixation tracks in order to avoid confounding automated and recallable cognition.

Suvorov (2015) also used eye tracking, in this case to understand video usage in second language listening test tasks. Language learners used the visual information to aid comprehension, and usage of videos differed according to how they were structured. However, to date there is no available research that uses of eye tracking to explore attention to feedback or processing of feedback from second or foreign language tests.

In addition, stimulated recall interviews have proven effective at eliciting evidence of a range of cognitive processes in second language learning. For example, Bao, Egi and Han (2011) studied learner noticing of recasts (where a fluent speaker correctly rephrases a learner's spoken error as part of natural conversation), and found that stimulated recall interviews with learners yielded higher rates of noticing than testing learners' ability to correctly reproduce the recasts. Egi (2008) found that stimulated recall interviews used in second language acquisition research did not appear to unduly impact participants' performance on posttest measures. Zhao (2010) investigated usage and understanding of peer and teacher feedback by English as a foreign language learners in China, and found that the interviews yielded useful information on learners'

understanding of feedback. Lam (2008) even advocated using stimulated recall in English as a second language classroom to provide a window on learners' metacognition.

Although an ethnographer of musicians, not language learners, Dempsey (2010) summarized the methodological benefits of stimulated recall interviews, noting that foremost, stimulated recall interview "can provide the ethnographer with details about the techniques people apply to successfully engage in a particular kind of activity" (p. 359). However, it should be noted that interview methods investigating cognition are limited by needing to infer processes based on reported thoughts, and processes that are fully automatized may not emerge at all in reported thoughts (Pintrich, 2004).

Many eye-tracking studies require no verbal contributions from participants. However, perhaps as a result of the noted limitations of each methodology, a number of studies have combined the use of eye-tracking traces with verbal reports of cognition from participants (Godfroid & Spino, 2015; Holmqvist et al., 2011). Sometimes these combined approaches involve using an eye tracking while the participant verbally reports their actions and cognition (cf. Smith, 2012); in other studies (cf. Godfroid & Schmidtke, 2013) the two methods are separated, with the verbalizations taking place after eye tracking has finished. Another methodological variation concerns the use of eye-tracking traces to prompt participant recollections. Bax (2013) used traces to prompt recollections, whereas Owen (2016) notes the benefits of first prompting without the use of traces, and then prompting again with the use of traces. In general, the relative benefits for each methodological choice depend on the research question at hand. However where, as in the present study, attention and cognition are of primary interest, it can be argued that triangulation of eye-tracking traces and participant verbalizations are essential, and that prompting free automatic recall is an important aspect of cognition to capture prior to exploring participants' interpretations of eye-tracking traces.

2.5.5 Summary

In summary, self-regulated learning is a key component of successful learning (Pintrich, 2004, Zimmerman, 2002), and a large part of self-regulated learning is cognitive and affective strategy use (Pintrich, 2004). While there is very little work investigating the cognitive and affective strategies that learners employ when negotiating language learning feedback, there is a substantial body of work that studies language learners' cognitive strategies while using a target

second language (cf. Phakiti, 2003; Raoofi, Chan, Mukundan, & Rashid, 2014; Sun, 2013), and there is also a substantial body of work that studies language learners' motivation and anxiety while learning or using a target second language (cf. Falout, Elwood, & Hood, 2009; Rose & Harbon, 2013; Swain, 2013; Yorke, 2011). Together with general research in cognitive psychology (Eisenberg, Sadovsky, & Spinrad, 2005; Monopoli & Kingston, 2012), these fields identify mental access to, then the appropriate selection and application of, cognitive and affective strategies – rather than simply applying more or a fixed set of strategies – as a major factor in academic and social success.

In the field of second language learning, a main implication of this finding is that lower proficiency language learners are unable to access the full set of strategies they possess in first languages due to the cognitive load of comprehending text (Lau & Chan, 2007; Lin & Yu, 2015; Saengpakdeejit & Intaraprasert, 2014). Other main findings relevant to processing and usage of language learning feedback are that affective strategies are closely related with attentional control (Wadlinger & Isaacowitz, 2011), and that motivation, beliefs about intelligence and goal orientation are all related with affective strategy use (Dörnyei, 2005; Garrett & Young, 2009; Mantou Lou & Noels, 2016).

Moreover, learners are generally positive about feedback received, although understanding whether learners comprehend feedback in the way intended is a tricky matter (Miller, 2009; Orsmond, Merry, & Reiling, 2005; Pitts, 2005; Walker, 2009). However, while relationships have been established between positive perceptions about feedback and usage of feedback (cf. Pokorny & Pickford, 2010; Poulos & Mahony, 2008; Timmers & Veldkamp, 2011), relationships appear influenced by multiple environmental and individual factors (Poulos & Mahony, 2008; Weaver, 2006), and research is clearly lacking on the ways in which these factors related to feedback acceptance and usage (cf. Furnborough & Truman, 2009; Jang, Dunlop, Park, & van der Boom, 2015).

Finally, eye-tracking and recall interview methodologies show promise for monitoring cognition of feedback on second language learning, as these methods have already been successfully applied to exploration of cognition in second language learning and second language learning (Bao, Egi, & Han, 2008; Egi, 2008; Kang, 2014; Bisson, van Heuven, Conklin, & Tunney, 2014 and second language testing (Bax, 2013; Ballard & Lee, 2015; Suvorov, 2015).

2.6 Individual characteristics that interact with language learning feedback usage

Throughout this literature review, the relevance of individual characteristics to processing and usage of feedback has been raised repeatedly. Indeed, Weir (2005) explicitly built test taker characteristics into his sociolinguistic framework for second language test validity due to the impact that such characteristics are known to have on test validity, so it follows that feedback from language test will also need to vary according to test taker characteristics. Already mentioned learner characteristics include second language proficiency, which is a known factor in the use of cognitive strategies (Lau & Chan, 2007; Lee-Thompson, 2008; Lin & Yu, 2015), while goal orientation and beliefs about intelligence have been observed to be related to differences in usage of affective strategies (Mantou Lou & Noels, 2016). In addition, the literature on assessment-for-learning and impact of feedback on academic achievement highlights the importance of language learners' goals and learning environment (Clark, 2012; Hattie & Timperly, 2007).

As learner characteristics clearly have a relationship with processing and usage of feedback, as well as feedback's utility for subsequent learning, and as the present study has adopted an effect-driven approach to feedback design, it can be argued that making feedback design choices based on a comprehensive understanding of the ways in which learner background characteristics interact with feedback processing and usage is a necessity. Therefore, existing research on the key learner characteristics previously named, and their relationships with usage of second language feedback, is now reviewed. Gaps in this literature will highlight areas to which this study can contribute, so that feedback design choices can be better informed.

2.6.1 Second language learning environment, goals and motivations

The impact of the social context on language learning has received much attention in the past 70 years, and justifiably so given the strong link between language and identity (Dörnyei, 2005; Ushioda, 2011). The role of identity interacts with language learning environment, producing language learning goals and motivations for learning that language. This study uses the L2 motivational self system (Dörnyei, 2005, 2009) theory of language learning motivation to understand these relationships. The system consists of three components: the ideal L2 self, the ought-to L2 self, and the L2 learning experience. The ideal L2 self is who the learner would like to become as an L2 speaker, the ought-to L2 self accounts for external pressures such as meeting

others' expectations and avoiding negative outcomes, and the L2 learning experience represents the language learning environment, over which learners may have varying degrees of control. Note also that language status and the instrumental utility of the target language forms part of the L2 learning experience (Clément, Baker, & MacIntyre, 2003; Csizér, & Dörnyei, 2005). Bearing these aspects of second language learning in mind, the environment and identities of adult immigrant English language learners are now considered.

This study provided feedback to adult immigrant English language learners in Canada. In Canada, immigrant language learners are known to struggle financially and with underemployment (Statistics Canada, 2009), and also with finding time to learn the sociocultural 'soft' skills that will enable them to integrate into Canadian society (Derwing & Waugh, 2012). However, immigrants themselves are highly motivated to achieve membership in Canadian society (Derwing & Waugh, 2012; Duguay, 2012; Han, 2009).

As a result, adult immigrant language learners hold powerful ideal English (as an L2) selves that may include integration with mainstream Canadian public and social life, white collar or skilled blue collar professional participation in the workforce, and strong relationships with English speaking family members such as children and grandchildren. Indeed, Ushioda (2011) argues that language learning motivation is unlike other learning motivations based on the

fact that language is a medium of *self*-expression and a means of communicating, constructing and negotiating who we are and how we relate to the world around us – that is, of giving ourselves voice and identity. A foreign language is not simply something to add to our repertoire of skills, but a personalized tool that enables us to expand and express our identity or sense of self in new and interesting ways and with new kinds of people; to participate in a more diverse range of contexts and communities and so broaden our experiences and horizons; and to access and share new and alternative sources of information, entertainment or material that we need, value or enjoy. (p. 204)

Yet as noted, immigrants can face language learning environments which are rich but in many cases restricted by child care responsibilities, limited access to the labour market, limited finances, and ghettoization. Cervatiuc (2009) showed that in these situations language learners are highly aware of the status of the target language and the personal stakes involved, and they negotiate their identities in this new situation through self-motivating dialogue, finding entry

points into native speaker communities, and building identities as ‘successful multilingual and bicultural adult immigrants’.

This relationship of the ideal L2 self with motivation and learning goals is consistently found in language learning contexts. Kormos, Kiddle and Csizér (2011) found that young adults in a Peruvian EFL context were motivated primarily by images of their future selves, and the strength of motivation depended on how close the individual felt to the international English speaking community, and by the cultural norms promulgated by guides such as parents and teachers. Similarly, Gu and Cheung (2016) found that South Asian students learning Chinese in Hong Kong was also determined by the ideal L2 self, in this case in terms of the extent to which the young adults desired to acculturate into mainstream Hong Kong culture – and once again, parental guidance impacted this desire. Wildsmith-Cromarty and Conduah (2015) likewise found that African immigrants to South Africa held similar motivations in learning indigenous African languages.

It is clear therefore, that status, utility and desire for shared identities play strong roles in motivating language learning – perhaps among the strongest roles. Moreover, a learners’ present and future environment is critical to mediating the strength and persistence of motivation. For example, Awad (2014) noted that positive experiences with language instructors, using the target language in real contexts, and family support impacted learner motivation to persist. Likewise, Matsumoto (2009) found that Australian young adult learners who were motivated to learn and associated themselves with Japanese culture were more likely to persist in learning Japanese. However, there is little information about how language learners’ motivations impact usage of feedback on language learning. The only study found, by DePasque and Tricomi (2015), indicated that stronger motivations for acquiring language were shown to be related to increased persistence in thinking about feedback when learners were tired or bored.

2.6.2 Goal orientation & beliefs about intelligence

Goal orientation can be defined as the reasons that individuals attempt to learn during a task. Two to four broad goal orientations are now widely documented in the literature. The goal orientations identified in this study are mastery, performance-prove and performance-avoid goal orientations (Button, Mathieu, & Zajac, 1996; Elliot & Harackiewicz, 1996; Midgley et al., 1998). Mastery-oriented individuals are motivated to master material, whereas performance

oriented individuals aim to create a positive impression. Performance-prove learners desire to prove they are competent to others, while performance-avoid learners attempt to avoid demonstrating incompetence to others. Multiple orientations coexist within a learner at any one time (Button, Mathieu, & Zajac, 1996; Jang, Dunlop, Park, & van der Boom, 2015), and mastery and performance prove orientations are associated with mixed learning outcomes (Elliot & Church, 1997). Performance avoid orientations are typically associated with lower academic achievement (Payne, Youngcourt, & Beaubien, 2007).

Type of sought feedback also differs by individual goal orientation. Individuals with dominant performance-goal orientations are likely to primarily seek positive feedback from high status authorities, whereas individuals with high mastery goal orientations are likely to seek accuracy feedback from experts (VandeWalle, 2003). Similarly, young learners with mastery orientations tend to disagree with their feedback and seek ways to improve their skills after receiving a feedback report, whereas those with performance prove orientations focus attention on their performance outcomes and tend to accept skill estimates (Jang, Dunlop, Park, & van der Boom, 2015). Tuckey, Brewer and Williamson (2002) found positive and negative correlations respectively between the desire for useful information, and mastery and performance goal orientations. Similarly, Pappachan (2008) found that mastery goal oriented employees valued positive and negative process feedback whereas performance goal-oriented employees sought only positive outcome feedback.

Moreover, goal orientation and associated feedback-seeking behaviour has been shown to be malleable and interact with skill level, personal stakes and other factors that change depending on context (Corpus, McClintic-Gilbert, & Hayenga, 2009; Kinlaw & Kurtz-Costes, 2007). In fact, Butler (1992) found that even artificially-induced mastery goal-oriented students sought more process feedback than outcome feedback. Butler (1993) also found that low-skill, performance goal-oriented participants sought much less computer-generated feedback than participants in other groups. Butler hypothesized that this was because these participants perceived a negative cost to their ego from viewing feedback. Vandewalle (2003) surveyed feedback-seeking literature and concluded that performance goal-oriented individuals are less likely to seek feedback than are individuals with learning goal orientations, particularly where poor performance and risks to self-efficacy and public image occur simultaneously.

Little research has been done specifically on the impact of goal orientation on language learning, but the work of Mantou Lou and Noels (2016), Gorges, Kandler and Bohner (2012) and Nakayama, Heffernan, Matsumoto and Hiromori (2012) indicates that goal orientation affects learning behaviours in ways typical to other domains, and is similarly malleable. Researchers note the importance of recognising these effects and opportunities for change given the strong feelings that many learners have about their ability to learn languages, and the unavoidably social performance context of language use.

Concurrent with goal orientations are beliefs about intelligence, for which a continuum describing two types of beliefs is recognized: at one end, people holding fixed beliefs about intelligence – also known as entity theorists, and at the other end, people holding incremental beliefs about intelligence – also known as incremental theorists (Dweck, Chiu, & Hong, 1995). Individuals holding fixed beliefs about intelligence believe that intelligence is innate, static and unchangeable; either one ‘has it’ or one does not. Individuals holding incremental beliefs about intelligence believe that intelligence is malleable and changeable; it can grow. Like goal orientations, beliefs about intelligence are known to be malleable (Aronson, Fried, & Good, 2002; Bonne & Johnston, 2016; Kinlaw & Kurtz-Costes, 2007) and have been shown to interact with underlying psychological characteristics such goal orientation, motivation, and self efficacy (Jang, Dunlop, Park, & van der Boom, 2015). For example, Thompson and Musket (2005) showed that goal orientation environment increases persistence among students with strong fixed beliefs about intelligence, while Stump, Husman and Corby (2014) and Davis, Burnett, Allison and Stone (2010) found that incremental beliefs about intelligence interact with self efficacy to relate to greater amounts of active, non-helpless learning engagement.

Beliefs about intelligence are an important construct to consider when developing feedback because beliefs about intelligence have been shown to relate primarily with willingness to persist in the face of difficulty (c.f. Yeager & Dweck, 2012) and acceptance of criticism (c.f. Mangels, Butterfield, Lamb, Good, & Dweck, 2006; Zhao, Zhang, & Vance, 2013), both of which are integral aspects of feedback delivery. For example, Zhao, Zhang and Vance (2013) demonstrated that students with fixed beliefs about intelligence were likely to self-handicap their performance on an assessment when they knew feedback was immediate, compared to those with incremental beliefs – whose performance significantly improved when they expected immediate feedback.

Likewise, Mangels et al. (2006) used biological indicators of cognition to show that learners holding fixed beliefs about intelligence engage in limited processing following negative feedback, leading to lower error correction.

For language learning, beliefs about ‘intelligence’ possibly represent beliefs about ‘language aptitude’, defined by Dornyei and Skehan (2003) as a “specific talent for learning foreign languages which exhibits considerable variation between learners” (p. 590). Notably, language aptitude is actually a known predictor of language learning success (Caroll, 1993), but it is not the only or even dominant predictor; motivation matches measures of aptitude in predictive strength (Dornyei & Skehan, 2003). Moreover, the two constructs are not significantly correlated (Dornyei & Skehan, 2003), indicating that the will to learn is at least as important as natural talent. These findings match research on beliefs about intelligence (discussed above), which consistently demonstrates that while variation in natural talent certainly exists among individuals, the belief that one can improve one’s intelligence (or perhaps aptitude) is positively related to behaviour that supports learning, and therefore in the long term becomes associated with success and achievement.

However as with goal orientation, little research on beliefs about intelligence has been conducted specifically with language learners, although language learning is a long term endeavor usually requiring substantial and sustained effort. Mantou Lou and Noels (2016) showed that language learners with fixed beliefs about intelligence were more likely to engage in helpless responses to feedback. Miele, Finn and Molden (2011) found that in vocabulary learning, language learners holding incremental beliefs about intelligence reported engagement with learning when doing effortful vocabulary learning, whereas those holding fixed beliefs were more likely to believe they would only recall learnt material if it were easy to learn.

2.6.3 Language proficiency

A key factor in usage of feedback for learning is the skill level of the learner. In the case of language learning, the issue is the language proficiency of the learner. Cognitive processing is substantially affected by specificity of information, but the effects differ according to skill level. In research on information grain size in instruction, a phenomenon known as the expertise reversal effect is well established (Kalyuga, Ayres, Chandler, & Sweller, 2003).

This effect is observed when different grain sizes benefit different expertise levels, and is theorized to be due to the existence of schema in higher-expertise individuals. Schema are 'chunks' of information that are stored as higher level cognitive processes in long-term memory, which are then accessed as single pieces of information, enabling higher expertise individuals to deal with more advanced information than novices, more easily.

For example, McNamara, Kintsch, Songer and Kintsch (1996) and Yeung, Jin and Sweller (1998) found that additions to an original text meant to increase coherence only benefited low-expertise students. Similarly, Kalyuga, Chandler, and Sweller (1998) found that experienced electrical students comprehended a circuit diagram better without additional text, while less experienced students benefited from the explanatory text. Worked examples in instructional context are another example of when increased detail tends to benefit low level learners but less so higher level learners (Kalyuga, Chandler, Tuovinen, & Sweller, 2001). The impact of adding detailed textual or visual information to language learning tasks and feedback has not been specifically researched.

In contrast, a substantial literature on the effects of hypertext on cognitive load finds that users of lower proficiency in the subject matter (Shin, Schallert, & Savenye, 1994), generally lower academic performance (Hailey & Hailey, 1998; Recker & Pirolli, 1995), and with smaller working memories (Lee & Tedder, 2003; Lin, 2003) achieve less well from hypertext learning material than from traditional linear text. Fontanini and Braga Tomitch (2009) replicated these findings with university students in the US who considered themselves English as a second language speakers. They found that lower language proficiency and lower working memory both inhibited use of hypertext. Chen and Yen (2013) made similar findings, observing that medium and high proficiency language learners gain vocabulary from digital annotations, but low proficiency participants did not. Moreover, Segalowitz and Frenkiel-Fishman (2005) present evidence that the ability of language learners to focus attention is determined by language proficiency, which has implications for delivering feedback in the second language.

These contradictory findings in separate circumstances indicate that different processes are occurring in the two contexts of apparent similarity in information density. An explanation is that lower proficiency learners, including language learners, need targeted information rather than

simply large amounts of information. Kalyuga (2009) offers a summary of recommended feedback during learning:

Detailed direct instructional support should be provided to novice learners as a substitute for missing knowledge structures in performing the executive role for high-level cognitive processes.... At higher levels of expertise, problem-solving or exploratory tasks that require using material available in long-term memory knowledge structures could be cognitively optimal instructional methods. (p. 5)

However, once again, no research was found looking specifically at the impact of second language proficiency on how language learners deal with feedback. This absence of research applied both to the relationship between language proficiency and dealing with feedback in the target language, and dealing with content density and type of feedback.

2.6.4 Self-assessment of language proficiency

Self-assessment potentially interacts with how learners use feedback because when results about skill levels have not matched learners' expectations, they may be more or less willing to accept feedback and associated direction about proposed learning paths. Yet while accurate self-assessment is considered a key aspect of self-monitoring progress and self-regulation of learning activities (Pintrich, 2004), Schlosser, Dunning, Johnson and Kruger (2013) note that the Dunning Kruger effect identified that in all areas of human knowledge and skill, individuals with low knowledge or skill levels substantially overestimate their abilities, while individuals with high knowledge or skill levels tend to slightly underestimate their ability. In fact, the changing relationship between actual skill level and perceived skill level can be graphed in the same way across multiple domains of learning. Kruger and Dunning (1999) argue this phenomenon is due to individuals with low skill levels lacking both the ability to demonstrate knowledge (because they do not have it), and the awareness to judge their ability because they do not know what high skill levels look like.

Indeed, this finding has been found repeatedly among language learners (Brantmeier, Vanderplank, & Strube, 2012; Jang, Dunlop, Park, & van der Boom, 2015; Ross, 1998; Trofimovich, Isaacs, Kennedy, Saito, & Crowther, 2016). As a reaction to this phenomenon, Sitzmann, Ely, Brown and Bauer (2010) argue through a meta-analysis of self-assessment research that self-assessment should be viewed more as an affective measure than a cognitive

measure, given its strong relationship with motivation and satisfaction, and moderate relationship with achievement. However, Richardson, Maeda and Swan (2010) argue in response that while self-assessment cannot replace external assessment, it has a valuable role to play in promoting and monitoring engagement with learning, and therefore has indirect effects on achievement.

This perspective is supported by Yoon and Lee's (2013) findings that using self-assessment over time had positive impacts on young English learner motivation and confidence, and in decreasing learner anxiety. Hirschel, Yamamoto and Lee (2012) also made similar findings for Japanese university students learning English, while Baleghizadeh and Masoun (2013) found increases in self-efficacy among Iranian young adults who regularly engaged in an English language learning program that required self-assessment, compared to those that did not. Huang (2016) noted that when provided with previous recordings of their own speech and the chance to engage in structured reflection, university English language learners in Taiwan undertook detailed consideration of their perceived strengths and weaknesses, and extended the reflection to their wider lives and learning.

Meanwhile, Bandura (1989) defines self-efficacy as the self-belief that one is capable of succeeding in a task, and Black and Wiliam (1998) note that learners who are aware of learning goals and feel task achievement criteria is demystified are more engaged in learning. That the empirical research finds learner motivation, reflection and self-efficacy increases while anxiety decreases when language learners regularly self-assess fits with the theories of self-efficacy and learner engagement. Note however that Anam and Stracke (2016) argued that self-efficacy has a positive impact on confidence in applying self-regulated learning strategies; self-regulated learning, self-assessment and self-efficacy are evidently related constructs.

Finally, there is evidence that direct experience with language use impacts the accuracy in self-assessment of language proficiency. Suzuki (2015) found that advanced Japanese language learners in Japan who had more experience using Japanese in real life made more accurate self-assessments of their proficiency. From a more immediate perspective, Goto Butler and Lee (2006) found that self-assessments based on specific tasks were more accurate than general self-assessments. These findings might again relate to self-efficacy, in that real life experience in the self-assessed domain is likely to orient levels of self-efficacy closer to actual current ability levels.

However, there is little research on how language learners deal with feedback in which language test results explicitly contradict language learners' own beliefs about language proficiency, for example by placing test and self-assessment results side by side. Alderson, Brunfaut and Harding (2014) argue for the value of presenting self-assessment alongside external diagnosis, as both sources of information are useful – even necessary, to develop an action plan. Moreover, Jang, Dunlop, Park and van der Boom (2015) found that doing so prompted strong emotional reactions and “appeared to draw students’ attention effectively and served as a reference for interpreting reported skill mastery levels” (p. 371).

2.6.5 Summary

In summary, there is a solid literature examining the effects of motivation (Derwing & Waugh, 2012; Dörnyei, 2005, 2009; Wildsmith-Cromarty & Conduah, 2015), goal orientation and beliefs about intelligence (Button, Mathieu, & Zajac, 1996; Corpus, McClintic-Gilbert, & Hayenga, 2009; Jang, Dunlop, Park, & van der Boom, 2015; VandeWalle, 2003), and language proficiency (Saengpakdeejit & Intaraprasert, 2014; Lin & Yu, 2015) on usage of feedback for learning, although limited work exists on how these individual characteristics affect feedback on language learning (cf. Fernandez-Toro & Hurd, 2014; Jang, Dunlop, Park, & van der Boom, 2015). In addition, much less work is available regarding the cognitive and metacognitive processes that learners engage in while processing feedback, and very little is specific to language learners dealing with feedback on their second language skills. This scarcity of research applies both to feedback on external assessment and self-assessment. Moreover, little work has been done to draw out the multiple relationships among language learners’ background characteristics and how language learning feedback is used. However, the literature on language learning, educational feedback, and the psychology of learning does indicate various aspects of feedback that should be adopted in order to maximize utility and potential benefits of feedback. These aspects are discussed below.

2.7 Feedback design recommendations

As metacognition is key to the use and implementation of feedback as part of self-regulated learning, feedback should ideally support learners who have less developed metacognitive skills to engage in metacognition. Support may be differentiated as information on what the learner can currently do (Nicol & Macfarlane-Dick, 2006) and information on where the learner can go next

(and how) (Clark, 2012; Hattie & Timperley, 2007). For lower proficiency language learners, the language of the feedback might need to be in a language that the learners can comprehend, in image form, or in relatively simple language. Likewise, as affect and motivation are other key areas within self-regulated learning, feedback should ideally assist learners in maintaining appropriate affect and motivation, as well as the related domain of self-efficacy. Design features that assist in focusing attention, such that appropriate orientations in these domains are maintained, may be particularly helpful.

Hattie and Timperley (2007), although not focused on second language education, developed a comprehensive model of feedback based on a review of research on educational feedback. Hattie and Timperley summarize their model by stating that feedback should “reduce the gap between current and desired understanding” (p. 86), and must answer three questions for the learner: “Where am I going? (What are the goals?), How am I going? (What progress is being made toward the goal?), and Where to next? (What activities need to be undertaken to make better progress?)” (p. 86). They argue that each of these questions needs to be addressed in terms of task comprehension and performance skills, self-regulation activities, and psycho-emotional factors.

In addition, Hattie and Timperley draw on existing research to highlight particular design features of feedback. Their intended context is the classroom, but several of the features are also relevant for the standalone nature of this study’s feedback. First, Hattie and Timperley highlight the importance of supporting goal setting and providing success criteria that helps learners understand what success looks like. Second they posit that feedback should inform learners regarding progress toward that goal. Finally, feedback should include tools, or direction to tools, that can assist learners in moving forward toward said goals. They note the importance of an appropriate amount of information and creating an environment wherein the learner is focused on learning. They also strongly argue for explicitly supporting self-regulation of learning through managing learners’ affect, promoting effective help-seeking behaviours, and creating opportunities to practice self-regulation, for example through supported goal setting and planning.

In terms of specific design features that might promote these positive effects, Goodman, Wood and Chen (2011) found that providing more specific feedback in training improved training performance, but reduced transferability of the skills to other situations. They argue that during

learning tasks, providing feedback of mixed specificity, as well as developing metacognitive problem solving skills, is likely to result in the strongest long-term training outcomes. Also on learning tasks, Chiviawsky and Wulf (2002) found that adults learned better when they were able to control when to ask for feedback. Moreover at a most basic level, the empirical quality of information provided in the feedback is essential to facilitating continued learning; inaccurate or inconsistent information (Chapelle & Chung, 2010) sends learners in unproductive learning directions or leaves them in confusion or indignation (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Walker, 2009).

It should also be noted that to date, little attention has been paid to how learners differ in their misunderstandings of feedback. In fact, intelligibility of educational feedback (cf. Carless, 2006) is a fairly new area of research itself. No research on the topic was found specific to language learning. However Higgins (2000) suggests that the most significant reason why students underutilize written feedback is that students simply cannot understand the feedback in order to apply it. This claim has some backing; Walker (2009) found that a large proportion of written comments provided in a distance learning information technology course were not understood by students, and Chanock (2000) found that almost half of humanities students interpreted a common written comment differently than instructors intended. Lea and Street (2000) also found that assumptions held by learners were often different than those held by instructors. Referring to the importance of integration of assessment and learning, Cramp (2011) found that face-to-face meetings to discuss written feedback improved university student perceptions of feedback intelligibility.

Finally, research is clear that the extent to which educational feedback is integrated into teaching and learning strongly impacts the effectiveness of that feedback (Alderson, Brunfaut, & Harding, 2014; Clark, 2012; Hattie & Timperley, 2007; McLaren, 2012; Sadler, 1989; Shute, 2008; Wiliam, Lee, Harrison, & Black, 2004). Without adequate and appropriate integration, feedback generally remains underutilized (Jang, 2005; Sadler, 1989). Recent studies report examples of how feedback can be integrated into programs of learning in innovative ways (e.g. Crisp & Ward, 2008; McLaren, 2012; Wang, 2007; Wiliam, Lee, Harrison, & Black, 2004). It is also at this point that computer-based environments may prove most useful, by offering time savings to teachers and learners regarding generation and delivery of feedback.

Chapter 3: Methods

This chapter outlines how the study was conducted in response to these questions. The research design is first overviewed, including a review of the research questions, and a justification for the mixed methods approach adopted in this study. Next, the study context is described, including the circumstances of the population from which participants were drawn, the English language test used, and why it was relevant to this population. Next, the participants are described.

Participant recruitment is first addressed, including criteria for invitation to participate in the study, a summary of participant numbers, and how these participants were recruited. Participants were sought via two types of network, so the recruitment methods for each type of network are described separately. A basic description of participant characteristics is also provided, including basic demographics such as gender and age range, as well as details about language backgrounds. However as participant characteristics are a major factor in different usage of feedback, participant characteristics are discussed in more detail in Chapter 4.

Next, the data collection procedures are described in detail, including the timeline and step-by-step activities. The instruments used in data collection activities are then described. A background survey is described, as well as how goal orientation and beliefs about intelligence composite variables were developed from the survey items. Next the Canadian English Language Proficiency Index Program (CELPIP) reading test is described as well as the process used to develop probabilities of mastery for each of the skills previously identified on the CDM. Following this, a self-assessment instrument is described, as is the process used to develop self-assessment estimates for each of the skills identified on the CDM. Next, the design choices for the report delivered to each of the participants are outlined and explained. Then, the design of eye-tracking and stimulated recall interview protocols are explained, followed by the design of a survey that participants completed after receiving their report and completed along with a description of the delayed recall interview protocol that was used with interview participants one month after receiving their reports. Finally, the analyses conducted are outlined, and the analysis procedures are described, including development of composite variables for further analyses, coding and analysis of interview and written response data, and analysis of eye-tracking quantitative and qualitative data.

3.1 Research design overview

Based on the literature review, this study posed several research questions designed to investigate adult immigrant English language learners' interaction with and usage of computer-based feedback on their English language reading skills. The research questions were organized into four themes, each with two to seven sub-questions:

- Theme 1 – Characteristics of adult immigrant English language learners in Canada
 - What are their occupational contexts?
 - What are their English language environments?
 - What are their English language goals?
 - What are typical goal orientation profiles and beliefs about intelligence?
 - What are typical English reading proficiency profiles?
 - How do they perceive their English reading proficiency?
 - What relationships are observable between individual characteristics?
- Theme 2 – Relationship between attention to and processing of feedback
 - To which aspects of feedback reports do language learners report paying attention?
 - To which aspects of feedback reports are language learners observed to pay attention?
 - Which aspects of feedback reports can be recalled by language learners one month later?
- Theme 3 – Language learners' affective and cognitive interaction with report content
 - What affective and cognitive strategies do language learners report when receiving a feedback report?
 - What processing outcomes do language learners report after receiving a feedback report?
 - In what ways are language learners' processing outcomes related?
- Theme 4 – Usage of feedback
 - How do language learners use a feedback report to plan their learning?
 - What are the relationships between individual characteristics, attention, processing experience, and usage of feedback?

The study adopted a non-experimental research design in which both qualitative and quantitative data was sought. The non-experimental design was required because participants are predefined by the population available in English as a Second Language programs, and they all received the same intervention. Both qualitative and quantitative data were sought because qualitative data allowed an open and deep exploration of participants' engagement with their report, while quantitative data permitted exploration of trends and patterns across all participants that were hypothesized but otherwise unobservable (Creswell & Plano Clark, 2007). In addition to enabling the use of different discovery approaches, a study design to collect only one type of data would be unable to adequately address the research themes. Specifically, the lack of existing literature on cognitive processes by which language learners engage with computer-based feedback on target language skills required using qualitative methods to build understanding of these processes, and the lack of existing literature on the impact of psychological factors on language learners' interactions with language learning feedback required using quantitative methods to test in what ways known psychological constructs may be associated with feedback usage.

Multiple mixed methods research design templates have been developed and taken up by educational researchers, including convergent parallel, explanatory sequential, exploratory sequential, transformative, and multiphase designs (Caruth, 2013; Creswell, 2012). This study adopted a convergent parallel approach in which both data types were given equal weight, and data collection took place in several phases. The timeline, procedures and instruments used to collect these data will be described in detail in the following sections. However, it is important to note at this stage that each of the datasets contributed understandings of one or two of the research themes.

3.2 Study context

This study sought to explore how second language learners process and use feedback on their second language proficiency that is delivered by computer and not teacher-mediated. Participants in this study were adult immigrant English language learners in Canada. The majority (80%) of participants were in government-funded full time English study programs, while a minority (20%) were in English medium high school or university programs of study. Participant recruitment and participant characteristics are described in further detail below.

The English language proficiency test most relevant to this population is the CELPIP, a general proficiency test of English as a Second Language in the Canadian context that is one of two tests accepted by the Canadian government for immigration purposes, and is also aligned with the Canadian Language Benchmarks, “a descriptive scale of language ability in English as a Second Language (ESL) written as 12 benchmarks or reference points along a continuum from basic to advanced” (Centre for Canadian Language Benchmarks, 2012, p. V) that is widely used among English as a Second Language education practitioners in Canada to establish a common theoretical grounding for instruction and to help teachers and learners plan “language learning, set or adjust goals, and monitor progress” (p. V), as well as to “facilitate a common understanding of learner credentials” across Canada (p. V).

Although the test developer and administrator, Paragon, does not currently release public information about the reading construct tested, a study on two CELPIP reading forms by Wagner, Chen, Park, Stone, and Jang (2015) identified typically recognized second language reading subskills (Grabe, 2009; Koda, 2007). Therefore, to provide feedback to adult immigrant English language learners in Canada on their English language proficiency, the present study utilized a cognitive diagnostic model (CDM) that had previously been developed on a retired CELPIP reading test form for Paragon Testing Enterprises by external researchers (Wagner, Chen, Park, Stone, & Jang, 2015). The information derived from the CDM is described in further detail below.

As a high-stakes language proficiency test for Canadian immigration purposes, constructive feedback from the CELPIP was of substantial interest to adult immigrant English language learners in Canada. Test takers were likely to want feedback to understand their relative strengths and weaknesses for making further progress toward proficiency goals. As a result, this context offered an authentic opportunity to research how language learners process and use language proficiency feedback in genuine language learning contexts where the test is not part of the learner’s regular program of instruction.

3.3 Participants

3.3.1 Participant recruitment

3.3.1.1 *Criteria for recruitment to study*

This study explored feedback processing among second language learners, and as noted, the test on which the feedback report was based (the CELPIP test, to be discussed below) was a test of English language proficiency for immigration purposes in Canada. Therefore, participants needed to be people who considered themselves non-native speakers of English, who intended to live and work in Canada over the long term, and ideally currently did not have permanent residence status. However, locating sufficient numbers of immigrants on work visas was logistically impossible, so in order to approximate the target population as closely as possible, adults who arrived in Canada within the last year were invited to participate. In addition, participants were required to consider themselves non-native speakers of English, to consider themselves to still be learning English, and to be intending to stay in Canada long term. Participants were sought via two methods: English programs for recent immigrants, and word-of-mouth. Details of participant recruitment via these two methods are now described.

3.3.1.2 *Participation summary*

While 111 people initially consented to participate in the study, 102 eventually participated in the study, with 98 people contributing a full basic dataset of background survey, test and self-assessment results, completed plans, and the report survey (see Table 5). Among the 102 participants, 15 people also additionally completed an eye-tracking interview, and 15 people completed a delayed recall interview one month later. A detailed exploration of participant profiles in terms of occupational situations and language use, psychological characteristics and language proficiency, is presented in Chapter 4.

Interview participants were selected based on logistical ability to conduct interviews and participant willingness to participate, and therefore the two sets of 15 participants are not overtly connected, although five participants completed both eye tracking and recall interviews. Note also that due to logistical limitations, only 15 interviews were able to be conducted for each set of interviews. Therefore the interview participants represent, out of 102 total participants, all the participants who were both able and willing to participate in an interview, rather than a subsample selected based on theoretical principles.

Due to the small number of interview participants, interview participant profiles are not reported in order to avoid facilitating either generalisations or claims of representativeness. However, it is worth noting that interview participation was primarily restricted by the logistical constraints of the study (e.g., equipment availability, time restraints), rather than the willingness of participants. An implication of this limitation is that interview participants are less likely to represent a particular subsample of the population.

Table 5

Total participants by data source

Data source	Number of participants
Full basic data set	98
Background survey	101
Test results	102
Self-assessment	101
Plans completed	101
Eye-tracking interview	15
Report survey (first page)	100
Report survey (second page)	101
Report survey (third page)	100
Delayed recall interview	15

3.3.1.3 Participant recruitment via English programs for recent immigrants

In Canada, the federal government funds English and French study for recent immigrants in order to promote integration into Canadian society. The English program is called the Language Instruction for Newcomers to Canada (LINC) program, and supports immigrants with rights to reside and work in Canada – broadly speaking those with permanent residence status but not yet Canadian citizenship – in a program of full time English study until they reach level 8 on the Canadian Language Benchmark framework, equivalent to approximately B2 on the Common European Framework of Reference (CIC, 2015; IELTS, 2017). The LINC program is implemented by language schools in colleges, community agencies, and immigrant settlement centres across Canada via contracts with the federal government (Burnaby, 2008; Fleming, 2007).

Five LINC programs based in Toronto, Canada consented to allow students in their intermediate classes to be invited to participate in the study. The five programs consisted of seven classes, and were located throughout the inner and outer areas of the city. People believed to be at Canadian Language Benchmark levels 5 to 8 were invited to participate, although some program students with level 4 accreditation were aware of the study and also requested participation; these people were allowed to participate as long as they were aware the test might be too difficult and likely not provide informative information for them. In total, 80 participants came to the study via this program-based route (see Table 6). Among these 80 participants, 13 participants from three classes completed eye tracking interviews, and 15 participants from four classes completed delayed recall interviews.

Table 6

Number of participants by participating school

Program	Total number of participants	Number of eye tracking interview participants	Number of delayed recall interview participants
School A	8	8	4
School B Location 1	14	0	4
School B Location 2	12	3	3
School B Location 3	9	2	4
School C	5	0	0
School D	15	0	0
School E	16	0	0

3.3.1.4 Participant recruitment via word-of-mouth networks

In addition, eligible individuals were contacted via word-of-mouth and invited to participate. These participants consisted of LINC program participants based in Kingston, Ontario, and international high school, undergraduate and graduate students based in Toronto. A further 22 participants came to the study via this word-of-mouth route (see Table 7). Among these participants, two participants from one network completed eye tracking interviews.

Table 7

Number of participants by word-of-mouth network

Community	Total number of participants	Number of eye tracking interview participants	Number of delayed recall interview participants
Word-of-mouth network 1	6	0	0
Word-of-mouth network 2	5	0	0
Word-of-mouth network 3	11	2	0

3.3.2 Participant characteristics**3.3.2.1 Basic demographics**

Of the 102 participants, three quarters (74%) of participants were female, and the median age was 38, with participant ages ranging from 17 to 69. As would be expected given Canadian immigration policy, which prioritizes individuals with high levels of education, two thirds of the participants (69%) held an undergraduate degree and one third (36%) also held a graduate degree. Most participants (79%) had not completed any of their education in Canada.

3.3.2.2 Language background

Reflecting general Canadian immigration trends, which includes immigration from all regions of the world, the 102 participants reported knowledge of 31 languages (including English) and reported 24 first languages. The largest first language groups were speakers of Chinese languages (39%), followed by Farsi (14%) and Spanish (13%) speakers. Languages of the Indian subcontinent represented 9% of first language speakers. This distribution is representative when use of English in countries of origin is considered (Statistics Canada, 2012); although Filipino and Indian sub-continent immigration to Canada is very high, these countries also have relatively high rates of English proficiency in the general population and therefore less need for English language support upon arrival, in contrast to China, Iran and Latin America, which are also regions from which many Canadian immigrants originate.

Note that 28% of these participants reported using English to communicate at home in some form, and 18% of participants reported being at least trilingual (at least two languages plus English). Therefore, many immigrants bring with them to Canada previous language learning experiences. Finally, also note that given the range of reported known and first languages, and

the complexity of identifying proficiency levels for reported languages, plus the complexities of identifying the impact of previous second or foreign language learning experiences on current second language learning, potential relationships between reported known languages and learners' interaction with and usage of their reports were not investigated in depth in this study.

3.4 Data collection procedures

Data collection took place between January and October 2015. An outline of the study's data collection activities is presented in Figure 1.

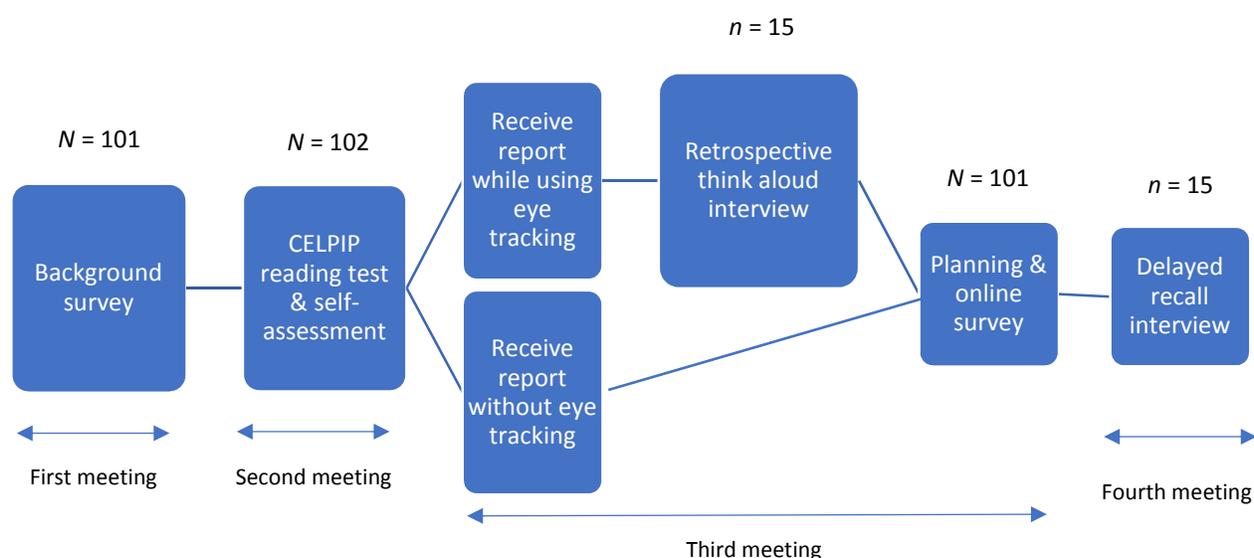


Figure 1. Data collection activities and order

At each location, data collection was initiated and completed within two weeks. At the first meeting, potential participants were introduced to the study and invited to participate, and they asked questions about the research. Potential participants then completed a consent form indicating whether they wished to participate and in which parts of the study, and those who wished to participate also completed a background survey that asked them about their demographics, English language use, and psychological influences (see Instruments section).

At the second meeting, participants did the CELPIP Reading test (see Instruments section) in groups of eight, and then immediately after completing the test also completed a self-assessment based on their perceptions of their performance on the test. The CELPIP Reading test was

administered using the standard test delivery system utilized for the CELPIP on laptops at the school or community location where participants were invited to participate. Participants completed the test in groups of eight because there were eight laptops available for the study. The self-assessment was paper-based (See Instruments section).

A day or so later, participants received their feedback reports and completed the report planning section (see Instruments section), and completed a survey that aimed to explore their interactions with the report (see Instruments section). Participants then downloaded a PDF version of the report to keep. The report was based online, so a link was emailed to participants along with a password, and they then logged in online to access the report and survey. The report and associated data were located on the University of Toronto's servers. In total, 102 participants received their feedback report and completed at least some of the planning section (although they did not necessarily complete the two surveys). In theory, no further face-to-face meetings were necessary. However, the participants who participated via English program classes all accessed their reports in a third face-to-face meeting, to maximize participation and to ensure smooth access to the report and usage of the survey. Participants who participated via word-of-mouth invitation accessed their reports online without a further face-to-face meeting.

An exception to this third step were the 15 participants who received their report while using an eye-tracking device. The main purpose of the eye-tracking device was to empirically track participants' gaze in order to explore attention among second language learners while receiving a feedback report on their second language proficiency. These participants all met face-to-face with the researcher a third time and received their reports individually using a computer that was integrated with eye-tracking technology.

In order to further delve into attention, and additionally to explore cognitive and affective processing strategies, the 15 eye-tracking participants also participated in a stimulated recall interview immediately after eye-tracking took place, in which participants first talked through their recollections of what they were thinking during their first reading of their report, then the eye-tracking traces were used to prompt further cognition recall. These participants were also asked three interview questions regarding their reflection on the report. After completing the interview, these participants then also completed the post-report survey online and downloaded their report to keep.

For most participants, this was the conclusion of their participation in the study. However, 15 participants also met individually one further time, a month after receiving their reports. These participants participated in a delayed recall interview that sought to identify the extent of participants' recall of the report content, and to explore their usage of the report since receiving it. Understanding the ability of participants to report aspects of the report after a month's time delay was important for exploring participants' attention. Understanding participants' usage of the reports informed how second language learners might be expected to actually use second language proficiency feedback reports. Therefore without prior reminding of their report content, participants were asked six questions designed to stimulate recall and prompted to elaborate.

3.5 Instruments

3.5.1 First meeting: Background survey

The background survey had two sections (see Appendix 1). The first section gathered data about demographics such as gender, age, language backgrounds, educational experience and occupations, and data about daily oral and written English language usage.

The second part of the survey consisted of 25 items designed to identify participants' goal orientations and beliefs about intelligence. Participants were asked to respond to how true each item was for them on a 5-point (1 to 5) Likert-style scale consisting of 'Not at all true', 'A little true', 'Somewhat true', 'Fairly true', and 'Very true'. For goal orientations, three goal orientations were measured: mastery, performance prove, and performance avoid (cf. Midgley et al., 1998). Each goal orientation was associated with seven items that were developed based on the Manual for Patterns of Adaptive Learning Scales (Midgley et al., 2000) and the goal orientation survey instrument used in the study reported in Jang, Dunlop, Park and van der Boom (2015). For beliefs about intelligence, two beliefs were measured: incremental beliefs about intelligence and fixed beliefs about intelligence (cf. Dweck, Chiu, & Hong, 1995). Each belief was associated with two items in order to keep survey length non-onerous.

3.5.2 Second meeting: Assessments

3.5.2.1 CELPIP reading test

The reading test from which the estimates of English reading proficiency and skills mastery estimates were derived was the CELPIP test. The CELPIP is a test of general English language proficiency test developed by Paragon Testing Enterprises. It is one of the two tests of general

English language proficiency that are currently accepted by the Canadian federal government for immigration purposes in Canada. It is also aligned with the Canadian Language Benchmarks. The test is taken via computer-based format in a secure test centre, and the reading section is automatically scored.

The reading test form used in this study was a retired test form from a previously operational CELPIP test. The test consisted of four tasks that required reading several text types and responding to 38 multiple choice items that were binary scored correct/incorrect. Each task was timed, and test takers needed to respond to all items on that task within the time limit, as the test automatically moved to the next task upon the expiration of the time limit. Prior to writing the test, these details were clearly explained to the study participants, and the test was completed under standard exam conditions.

3.5.2.1.1 Development of probability of mastery estimates for reading skills

In order to provide substantive information for learning, this study implemented a cognitive diagnostic model (CDM) that had previously been developed for the reading test form used in the study to develop the test-based feedback. The pre-developed CDM used was a reduced Re-parameterized Unified Model (RUM) (Roussos et al., 2007) developed for the 38-item reading test, in which a Q-matrix of ten pre-identified skills was used (Wagner, Chen, Park, Stone, & Jang, 2015). Applying a pre-developed CDM to subsequent participants' test item responses produces estimated posterior probabilities of mastery (*PPMs*) for each skill identified in the CDM, for each participant. *PPMs* are the probability that an individual has mastered a specific skill, and are therefore represented on a scale of 0 to 1.

As grain size and amount of information is a known key area for creating increased overwhelmedness and/or sense of satisfaction (Kalyuga, Ayres, Chandler, & Sweller, 2003), all ten skills were not automatically included in the report. In this particular study, the six most frequently assessed skills (Wagner, Chen, Park, Stone, & Jang, 2015) were identified for inclusion on the report in order to allow learners to focus on a range of reading subskills, but hopefully without overwhelming them. The six skills identified for inclusion in the report were Using vocabulary knowledge (VOC), Comprehending explicit textual information (including skimming, scanning and matching) (EXP), Comprehending implicit meaning (IMP), Making

connections through inferencing (INF), Distinguishing main ideas from minor details (DIS), and Understanding discourse context and social situation (PRG) (see Table 8).

Table 8

Reading subskills included on the report

Reading subskill	Skill definition	Number of items	Mean <i>PPM</i> *
Process vocabulary knowledge independent of text (VOC)	Ability to process knowledge of word(s) independent of provided textual information to make meaning	16	.55
Comprehend explicit textual information skim, scan, match (EXP)	Ability to use strategies such as skimming, scanning, and matching to locate textually explicit information	12	.59
Comprehend implicit meaning (IMP)	Ability to process implied meaning in text by retelling or rephrasing	10	.37
Make connections through inferencing (INF)	Ability to infer meaning from text including: future events (predicting), author's intent, causal consequences	9	.45
Distinguish main ideas from minor details (DIS)	Ability to distinguish main/key ideas from details across sentences and paragraphs	9	.63
Understand discourse context and social situation (PRG)	Ability to identify the context or situation in which conversations take place and in which characters relate	7	.65

* Mean *PPM* for population on which the CDM was originally developed, $N = 796$ (Wagner, Chen, Park, Stone, & Jang, 2015).

The estimates of skill mastery were reported to participants on their feedback report. Participants' raw total test score was also used in subsequent analyses.

3.5.2.2 Self-assessment tool

The self-assessment tool from which self-assessment scores were derived for the report was developed based on the skill descriptors identified in collaboration with practicing LINC

program teachers for use in the report. As there were six skills on the report, each with three descriptors, the self-assessment therefore consisted of 18 items, with three items per skill (see Appendix 2). The descriptors were randomized and participants were asked to rate on a five-point Likert-style scale how often they found they could do these descriptors while doing the test. The scale consisted of ‘Never’, ‘Rarely’, ‘Sometimes’, ‘Usually’ and ‘Always’. Participants were reminded to answer the self-assessment items based on their experience in the test, not about their general daily experiences. This choice was made to maximize the similarity between the test scores and self-assessment scores. Research also indicates that self-assessments of language skills are more accurate when learners consider more specific language usage situations (Goto Butler & Lee, 2006).

3.5.2.2.1 Development of self-assessment scores for reading skills

Participants responses were coded on a scale of 1 to 5, with 1 representing ‘Never’ and 5 representing ‘Always’. For inclusion in the report, the mean of the three items associated with each skill was calculated and transformed into a 0 to 1 scale using the formula:

$$y = .25 (((x1 + x2 + x3) / 3) - 1)$$

The resulting scores were then delivered to participants on the report figures.

3.5.3 Third meeting: Receiving the feedback report

3.5.3.1 Feedback report

The report was designed to provide substantive, useful information for learning English, specifically for improving English reading skills. In order to meet this goal, the report consisted of several components. First, an introduction (see Figure 2) provided an orientation for the learner. The language of the introduction was reviewed and edited by four practicing LINC program teachers.



Figure 2. Example of introduction to report (fictional learner)

Learners were then presented with their test and self-assessment results. For each of these skills, the report gave a plain-language title for the skill, three bullet points listing ‘can do’ statements that described what the skill was, and a figure that showed the probability of mastery alongside the learner’s self-diagnosed probability of mastery. The self-diagnosed probability was derived from a self-assessment. The ‘can do’ statements were originally sourced from the Canadian Language Benchmark documents for levels 5 to 9, and four practicing LINC program teachers were invited as experts to identify which skill(s) each descriptor represented, and whether the activity was easy, at-level or difficult, for learners around Canadian Language Benchmark 8. Based on their feedback, three ‘can do’ statements were selected for each skill. The language for the statements, and the rest of this section, was then reviewed and edited by the same teachers. Figure 3 shows an example report of a fictional learner.

Your English reading skills

This report tells you about 6 reading skills. The information comes from your responses on the CELPIP reading test and your self-assessment. For each skill, the report tells you examples of how you use that skill. The report tells you how well you think you do this skill, and how well you did this skill on the test.

Are you surprised? Do you agree?

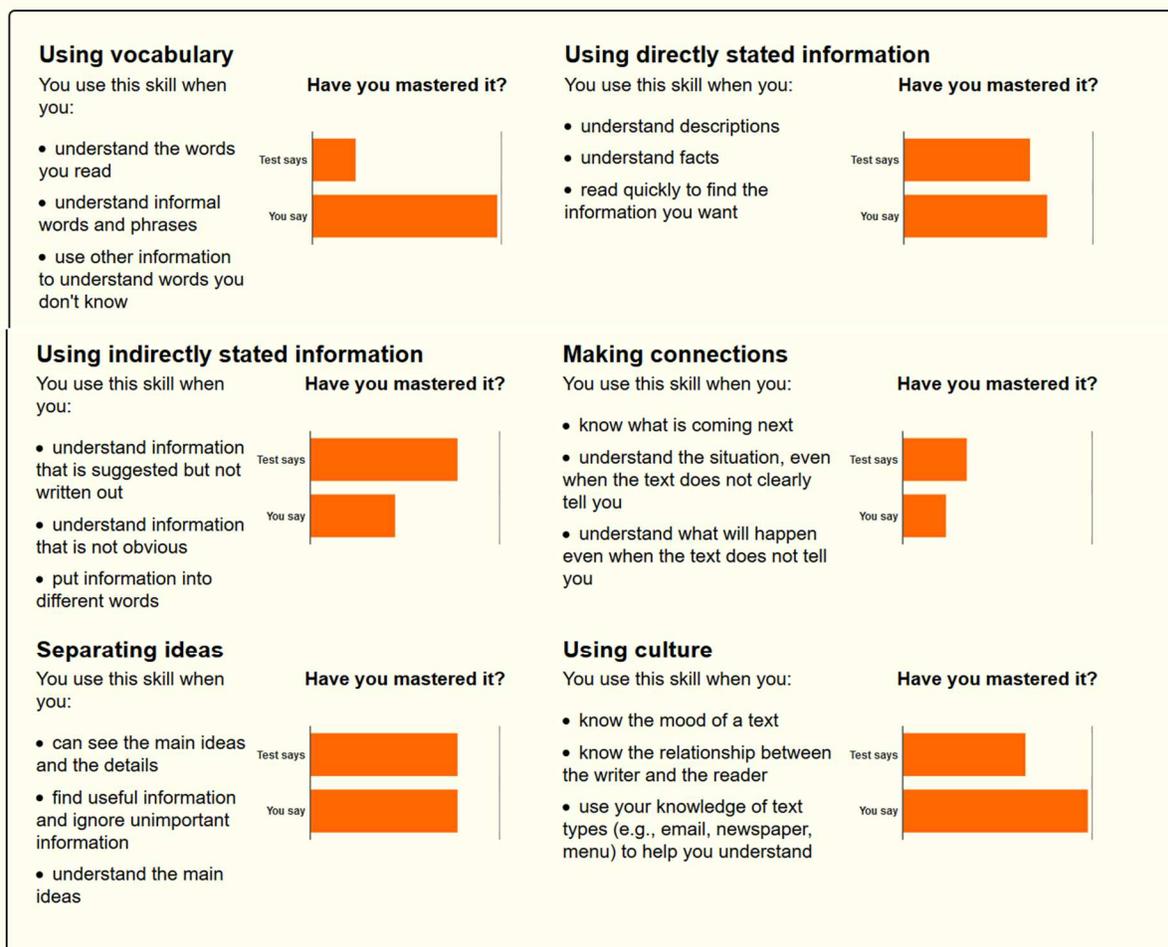


Figure 3. Example of report skill descriptions and learner personal results (fictional learner)

Following the skill descriptions and figures, the report provided some learning suggestions (see Figure 4). Each learner was suggested to work on two skills, and the suggestions varied according to the test-based skill profile of the individual. For each skill, three activities were suggested that would help the learner improve that skill. The suggested activities were developed in consultation with four practicing LINC program teachers. Activities were brainstormed with the teachers individually, and three for each skill were eventually selected in consultation with

the teachers. The teachers also reviewed and edited the language in the learning suggestions. An example of learning suggestions is in Figure 4 below, and the full list is available in Table 9.

Suggestions for learning

Here are some ideas for study. These activities may help you learn better English.

Learners with your reading skills profile often want to improve their ability to **use vocabulary**.

- Ideas that may help you do this are:
reading example sentences of words you are not sure about,
choosing a text you can more or less understand and checking the meanings you are not sure about, and
watching movies and TV with subtitles.

In addition, you may want to improve your ability to **make connections**.

- If so, try:
reading lots of information about similar topics,
predicting what you will read next, then seeing if you are right, and
thinking about and comparing what you know with what you read.

Do you agree? What do you want to work on? What will help you?

Figure 4. Example of report suggestions section (for skills VOC and INF)

Table 9

Report learning suggestions, by skill

Using vocabulary (VOC)

reading example sentences of words you are not sure about

choosing a text you can more or less understand and checking the meanings you are not sure about

watching movies and TV with subtitles

Using directly stated information (EXP)

reading labels in your house and on your food, and reading signs and notices on the street

reading and listening widely to material you can more or less understand

looking for chances to tell stories and give instructions and descriptions

Using indirectly stated information (IMP)

discussing with other people the information a writer wants to communicate

thinking about the writer's feelings while you are reading

thinking how what you have read could be said another way

Making connections (INF)

reading lots of information about similar topics

predicting what you will read next, then seeing if you are right

thinking about and comparing what you know with what you read

Separating ideas (DIS)

finding the important ideas in each paragraph

finding the main point of what the author wants to tell you

looking away from a text and thinking about the key ideas

Using cultural knowledge (PRG)

volunteering with local organizations and at events

listening to radio news and reading newspapers daily

taking part in Canadian holidays, festivals and hobbies

Finally, participants were required to engage in some planning for their learning. They were required to select between one and three of the skills that were presented in the report, and identify learning goals for that skill, action plans for how they would achieve their goals, and monitoring plans so they would know how they were making progress. The basis for this structure was work by Hattie and Timperly (2007) that indicates that learners need to know where they are going and how they are going to get there and additional work by Pintrich (2004) that indicates that monitoring of progress is key to keeping on target. The planning section introduction and the first plan outline are shown in Figure 5 below; the next two plan outlines are the same. Participants' input into this planning section was subsequently used in analyses; the analyses are described in the analysis section.

Plan your learning

Use the information in your report to set learning goals for yourself. Some questions will help you think about your learning goals.

- Choose 1-3 skills you want to work on this month. Choose skills from the report.
- For each skill, answer these questions:
 - What do you want to achieve? **Try to be clear.** You can use the skill examples to help you.
 - What will you do to achieve this goal? Write 1-3 ways you will work on this skill. **Try to be clear.** You can use the suggestions for learning to help you.
 - How will you **check your progress** toward your goals?

* Required

I want to work on *	Select...
My goal for this skill *	<input type="text"/>
What I will do to reach this goal *	<input type="text"/>
How I will check my progress *	<input type="text"/>
I also want to work on *	Select...
My goal for this skill	<input type="text"/>

Figure 5. Example of report planning section (first plan outline only)

3.5.3.2 Eye-tracking and stimulated recall interview protocol

For eye-tracking and stimulated recall interview activities, an eye-tracker was arranged and an eye-tracking/interview protocol was developed (see Appendix 3). The eye-tracking device used was a Tobii 2150 with a 21 inch monitor. The eye-tracking section of the protocol included introducing the device to participants, training the eye tracker, and delivering the report. The stimulated recall interview section of the protocol provided prompts for guiding participants' recall of cognition, first without using eye-tracking traces and then a second time with using them, and three interview questions designed to prompt participants to discuss their current reflections on the report content.

3.5.3.3 Post-report online survey

The post-report online survey explored participants' feedback processing outcomes through eliciting information on four general constructs: attention, feeling able to cope with amount of report content, cognitive and affective outcomes with respect to report content, and intent to use the report for further learning. These constructs were based on existing literature on cognition in feedback and learning, as covered in the literature review. Participants' planned usage of the

report was additionally further explored by asking them who they planned to talk to about the report, and what they planned to talk about. The set of survey items used in this study is provided in Appendix 4.

The construct of reported attention was addressed by asking participants how much they looked at each section of the report: introduction, skill descriptions, figures, suggestions, and planning. Participants responded on a Likert-style 1 to 5 scale of ‘I did not look at this’, ‘A little time’, ‘Some time’, ‘A lot of time’ and ‘All my time’, or could select ‘What is this?’.

To explore participants’ sense of being able to cope with the amount of information in the report, participants were asked to state how true eight items were for them on a Likert-style 1 to 5 scale of Not true -> -> -> True, plus an additional option of ‘I don’t understand [this item]’. The eight items addressed participants overall sense of the report, and were designed to access two hypothesized sub-constructs regarding ability to cope with the amount of information: sense of overwhelmedness and extent of thinking about the whole report.

To explore participants’ cognitive and affective outcomes with respect to the report content, participants were asked to state how true eight items were for them on a Likert-style 1 to 5 scale of Not true -> -> -> True, plus an additional option of ‘I don’t understand [this item]’. The eight items were designed to access two hypothesized sub-constructs of processing: trust in the report content and reflecting on one’s English skills as a result of reading the report.

To explore participants’ intentions regarding using of report for further learning, participants were asked to state how true eight items were for them on a Likert-style 1 to 5 scale of Not true -> -> -> True, plus an additional option of selecting ‘I don’t understand [this item]’. The eight items were designed to access two hypothesized sub-constructs regarding intended application: desire to use the report and sense of needing help to use the report.

In addition, participants were asked who they wanted to talk to about the report. They were asked to state how likely they were to talk to three groups of people about the report, on a Likert-style 1 to 5 scale of Very unlikely -> -> -> Very likely. The three groups of people were their English teacher, their English program classmates, and their friends and family outside English class.

Finally, participants were asked what they would probably discuss with these people. They were invited to select all the topics that applied to them, out of a total of seven topics: about how my

self-assessment compares to the test results, how much I agree or disagree with the report, how to improve my English, my level of English proficiency, what areas of my English skills I should focus on, what the other person thinks about my English skills, and what the report says about my English language skills.

3.5.4 Fourth meeting: Delayed recall interview protocol

A delayed recall interview protocol was developed to structure the recall interview that took place one month after participants received their report (see Appendix 5). The protocol was designed to be a semi-structured interview, with six questions to prompt recall of report usage and content:

1. How have you used your reading report?
2. What do you remember about your reading report?
3. What did the report tell you about your reading skills?
4. What did the report look like?
5. What did your report suggest you should do to continue learning?
6. What learning goals did you set?

3.6 Overview of data sources

In summary, the purpose of adopting a mixed methods approach while giving equal priority to various data types was to explore the research themes in different ways to develop rich responses to the research themes, and fuller understandings of phenomena underlying those responses.

Table 10 below summarizes how each dataset contributed to the research themes.

Table 10

Overview of data sources

Meeting	Activity (n)	Data	Purpose
First and second	Background survey, reading test and self-assessment (102)	Close-ended survey responses	Described participant background characteristics (Theme 1)
		Test results	Explored how participant characteristics impact
		Self-assessment results	feedback usage (Theme 4)

Meeting	Activity (n)	Data	Purpose
Third	Report planning section (102)	Planning section responses	Identified general trends and participant differences in using feedback for planning (Theme 4)
	Eye tracking (15)	Gaze paths Fixation times	Identified underlying attentional processes and priorities (Theme 2)
	Eye-tracking stimulated recall interview (15)	Retrospective think aloud transcripts	Identified underlying attentional processes and priorities (Theme 2) Explored cognitive and affective processes when dealing with feedback (Theme 3)
	Online survey (102)	Close-ended survey responses	Identified reported attention to feedback (Theme 2) Identified general trends and participant differences in processing outcomes and intended usage of feedback (Themes 3 and 4)
Fourth	Delayed recall interview (15)	Delayed recall transcripts	Explored usage and recall of feedback (Themes 2 and 4)

3.7 Analyses

Analysis took place in two main stages. The first stage was preparatory analysis, in which quantitative data was cleaned and composite variables were developed and validated, and qualitative data was coded iteratively into themes or variables, depending on the data source. Table 11 presents an overview of these preparatory analyses, and they are discussed in detail in the next section, organized by first, second, third and fourth meeting. The second stage of analysis was descriptive and inferential analyses relevant to specific research themes. Table 20 presents an overview of these analyses, and they are described in the section after preparatory analysis, organized by research theme one to four. Results of the theme-specific analyses are reported in the results section. Quantitative analyses were conducted using Stata IC version 14. Microsoft Excel was also used to support quantitative and qualitative analyses.

Table 11

Overview of preparatory data analysis

Meeting	Data	Preparatory analysis
First and second	Close-ended survey responses	Developed and validated composite psychological and self-assessment variables
	Test results	Comparison of CDM-derived <i>PPMs</i> and raw total score, validation of raw total score variable
	Self-assessment results	
Third	Planning section responses	Developed variables representing relevant characteristics of planning behaviour
	Gaze paths	None
	Fixation times	
	Retrospective think aloud transcripts	Developed codes representing observed attention, affect and cognition
	Close-ended survey responses	Developed and validated five composite variables of affect, cognition and intended use
Fourth	Delayed recall transcripts	Identified who used their report, and identified themes regarding why report used or not used

Table 12

Overview of theme-specific data analysis

Theme	Purpose	Theme-specific analysis
1	Described participant background characteristics	Descriptive and correlational analyses of demographic, English usage, psychological and language proficiency variables Three latent profile analyses: goal orientation variables, CDM-derived <i>PPMs</i> , and raw total score versus self-assessed score
2	Identified underlying attentional processes and priorities Explored recall and usage of feedback	Descriptive analysis of reported attention Descriptive analysis of fixations by report section Comparison of eye-tracking paths Exploration of which report sections were recalled, and associated themes

Theme	Purpose	Theme-specific analysis
3	Explored cognitive and affective processes when dealing with feedback Identified general trends and participant differences in processing outcomes and intended usage of feedback	Description of observed processes and exploration of their relationships Descriptive analyses of composite variables, followed by latent profile analysis of the variables
4	Identified general trends and participant differences in using feedback for planning	Descriptive analyses of planning variables Wilcoxon-Mann-Whitney tests comparing intended discussion partners Comparison of planned activities with activities done Exploration of who used their report and who did not, and reasons given Inferential analyses identifying background and processing variables related with usage of feedback: correlational analyses, a Kruskal-Wallis test, and linear, ordinal and logistic regression analyses

3.7.1 Preparatory analyses

3.7.1.1 *First meeting: Background survey*

The data from the goal orientation section of the background survey required condensing into theoretically homogenous composite variables before it could be used for analyses concerning the research themes. The survey contained items hypothesized to represent three goal orientations: mastery, performance prove, and performance avoid, plus several items hypothesized to represent beliefs about intelligence: fixed and incremental. The process by which composite variables were developed for these constructs is described below.

To develop the composite variables of each of the three hypothesized goal orientations, the correlations among items were considered (see Table 13 and Table 14). An exploratory factor analysis (Fabrigar, Wegener, MacCallum, & Strahan, 1999) was also conducted (see Appendix 6 Exploratory factor analyses), but due to the study sample size ($N=102$) being at the absolute lower end of acceptable sample size, a robust exploratory factor analysis was not possible (Fabrigar, Wegener, MacCallum, & Strahan, 1999), so final composite variables were not based on factor scores derived from these results. However, the exploratory factor analysis results were

similar to the factor structure observed in the simple correlation matrix, supporting the use of composite variables based on mean scores.

Most goal orientation items were positively associated with the other items in their construct, and not associated or negatively associated with other items. There were three exceptions. One item was intended to be a performance avoid item but correlated with performance prove items. Upon investigation the item ('Good grades are the most important thing for me', item PA5) was exposing performance prove goal orientations, and was moved into the performance prove construct. Another item weakly correlated across both performance prove and performance avoid items. The item ('If I have trouble doing a task, I don't tell anyone', item PA7) applies to both constructs and was omitted to promote clarity of constructs. A third item failed to correlate consistently with any construct. The item ('I only feel successful if other people tell me I did a task well', item PP5) was therefore omitted from the composite variables. The composite variables were created by calculating the mean score of all the items associated with each construct. Coefficient alphas for the composite variables were .76 for mastery goal orientation, .80 for performance prove goal orientation, and .69 for performance avoid goal orientation.

Table 13

Spearman's rho correlations among goal orientation items

	m1	m2	m3	m4	m5	m6	m7	pp1	pp2	pp3	pp4	pa5	pp6	pp7	pa1	pa2	pa3	pa4	pa6	pa7	pp5	
m1	1.00																					
m2	*.45	1.00																				
m3	*.58	*.27	1.00																			
m4	*.39	*.35	*.35	1.00																		
m5	*.26	.16	*.21	*.29	1.00																	
m6	*.58	*.42	*.63	*.35	*.43	1.00																
m7	*.41	*.41	*.40	*.49	*.25	*.40	1.00															
pp1	.04	-.06	-.11	.10	.18	.04	-.11	1.00														
pp2	.04	-.04	.12	.18	.11	.09	-.02	*.30	1.00													
pp3	-.03	*-.25	.00	-.12	-.00	-.14	-.14	.13	*.37	1.00												
pp4	-.03	-.05	.11	-.06	.11	.07	-.00	*.38	*.41	*.38	1.00											
pa5	-.03	.02	.01	.19	.18	-.03	.05	*.2	.16	*.37	*.50	1.00										
pp6	.01	-.03	-.01	.09	.04	.06	.04	*.42	*.46	.22	*.34	*.42	1.00									
pp7	-.06	-.03	.04	.15	.19	.01	.03	*.35	*.63	*.34	*.49	*.34	*.42	1.00								
pa1	*-.21	-.15	*-.24	*-.16	-.14	*-.42	-.19	.01	-.04	*.20	.07	.09	.02	.02	1.00							
pa2	.08	-.08	.09	-.12	*.21	.01	-.02	.11	*.23	*.29	*.38	*.27	-.03	*.21	*.26	1.00						
pa3	*-.26	-.17	*-.20	-.14	-.03	-.19	-.09	-.03	*.22	.17	.12	*.23	.19	.18	*.20	*.25	1.00					
pa4	.02	-.07	-.10	.15	-.09	-.11	-.05	.04	.16	.13	.06	*.19	.03	.19	*.33	*.43	*.25	1.00				
pa6	-.12	-.13	-.10	-.19	-.08	-.18	-.10	-.07	.11	.19	.17	.12	-.01	.15	*.38	*.39	*.39	*.48	1.00			
pa7	*-.21	-.09	-.06	-.15	.06	-.11	-.12	-.04	*.25	*.26	*.26	*.23	*.21	*.27	*.30	.06	*.29	.18	*.39	1.00		
pp5	.06	-.14	.03	.03	.03	-.15	.08	.11	.15	.06	.12	*.21	*.31	*.24	.10	-.00	.18	*.20	.13	.06	1.00	

* $p < .05$; $N = 97$ to 101 ; Spearman's ρ

Table 14

List of items with codes

Code	Item
m1	It's important to me that I learn new things when I do tasks
m2	I really want to understand what I am learning
m3	When I am doing tasks, I try to learn new skills
m4	I prefer challenging and difficult tasks so I can learn new things
m5	I choose challenging tasks even if I might not be very successful
m6	When I am doing tasks, I enjoy learning as much as I can
m7	It's important to me that I improve my skills when I do tasks
pp1	When I do tasks, I want to be more successful than most other people
pp2	I make sure other people know when I am successful on a task
pp3	When other people can do a task, I want them to know it is easy for me
pp4	It's important for me to do better than other people on tasks
pa5	Good grades are the most important thing for me
pp6	I want to do well because I want to show my ability to my family, friends, or teachers"
pp7	I really like to show other people that I can do tasks
pa1	I do not ask questions if I might look stupid
pa2	It's important to me that I don't look stupid on tasks
pa3	I prefer easier tasks that make me look good
pa4	I don't want to look stupid so I choose tasks I can do well
pa6	I stay away from tasks where other people might think I'm not smart
pa7	<i>If I have trouble doing a task, I don't tell anyone</i>
pp5	<i>I only feel successful if other people tell me I did a task well</i>

Similar to the goal orientation items, to develop the beliefs about intelligence composite variables, the correlations among the four relevant items were considered (see Table 15). In this case, the number of items was too small to conduct an exploratory factor analysis. The items for incremental beliefs about intelligence moderately correlated ($\rho = .42, N = 99, p < .000$), so the mean of these two items was calculated for the composite variable of incremental beliefs about intelligence. The items for fixed beliefs about intelligence did not significantly correlate ($\rho = .17, N = 96, p = .078$), so the two items were retained separately for inclusion in analyses individually, to see which item was operating as the most effective indicator of fixed beliefs about intelligence.

Table 15

Spearman's rho correlations among beliefs about intelligence items

	F1	F2	I3	I3
People can only learn languages well if they are born with language learning ability (F1)	1.00			
People are either good or bad at learning languages (F2)	.18	1.00		
If someone works hard, they can learn any language (I3)	-.06	.02	1.00	
With effort, any adult can learn a language fluently (I4)	-.18	-.08	***.42	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$; $N = 96$ to 100

Finally, Pearson's r correlations between the composite variables were calculated to verify whether the composite variables were related in the way that existing research indicates (see Table 16). Note that while raw items were considered ordinal scales given their 1-5 rating scale and therefore correlated with Spearman's ρ , the composite variables were correlated using Pearson's r because the composite variables, being mean scores, were best considered continuous variables. Mastery goal orientation was uncorrelated with performance prove ($r = .06$) or performance avoid orientations ($r = -.13$), but positively correlated with incremental beliefs about intelligence ($r = .22$). In contrast, performance avoid and performance prove orientations were related with each other ($r = .30$) and positively related with fixed beliefs about intelligence (e.g., F1 r s = .25 and .23). These relationships reflect the literature (Midgley et al., 1998) and confirm that the composite variables can be used to represent these constructs.

Table 16

Pearson's r correlations between goal orientation and beliefs about intelligence composite variables

	M	PP	PA	I	F1	F2
Mastery (M)	1.00					
Performance prove (PP)	.06	1.00				
Performance avoid (PA)	-.13	** .30	1.00			
Incremental beliefs (I)	* .22	* -.23	-.07	1.00		
People can only learn languages well if they are born with language learning ability (F1)	.03	* .25	* .23	-.08	1.00	
People are either good or bad at learning languages (F2)	-.06	.12	** .32	-.07	.17	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$; $N = 96$ to 101

3.7.1.2 Second meeting: Assessments

The reading test data and self-assessment data also needed some analyses in order to inform subsequent analyses. For the reading test, there were CDM-based probabilities of skill mastery for each of the six skills reported on the test, as well as a set of raw binary responses. The distribution of scores for each of these metrics was of interest to understand participant English reading proficiency. In addition, the relationship between these two metrics of English reading proficiency was compared by using the raw total of the binary responses – the raw test score, and the probabilities of skill mastery. Based on this comparison, key raw scores were identified that represented observable cutoffs in terms of probability of mastery. As a result, the raw total score could be used as a single score to represent English reading proficiency.

For the self-assessment, analyses were conducted to identify whether a single score could be developed for self-assessed English reading proficiency. The analyses indicated that it could, therefore this score was developed for use as a composite score of self-assessed English reading proficiency in subsequent analyses.

3.7.1.2.1 CELPIP reading test

Among the 102 participants, the estimates of skill mastery showed a binary distribution, which is typical for CDM estimates of skill mastery (Buck & Tatsuoka, 1998; Jang, 2005; Jang, Dunlop, Wagner, Kim, & Gu, 2013), although a higher-than-usual proportion of participants had intermediate probabilities, which is probably a result of the learner levels targeted – they were intended to be moving toward mastery of the tested subskills. The relatively harder skills were Using implicitly stated information (mean $PPM = .18$, $SD = .31$), Making connections (mean $PPM = .22$, $SD = .33$) and Using vocabulary (mean $PPM = .27$, $SD = .40$), while the relatively easier skills were Separating ideas (mean $PPM = .43$, $SD = .38$), Using culture (mean $PPM = .36$, $SD = .43$), and Using directly stated information (mean $PPM = .35$, $SD = .39$). The median raw total score was 16 out of 38 (42%) with a range of 4 to 35. Estimates of skill mastery probabilities and total scores correlated moderately to strong, ranging from $r = .67$ to $.89$ ($N = 102$, $p < .001$). A full set of correlations is presented in Table 17. Therefore, the raw total score was used to indicate English reading proficiency where a single score was necessary.

Table 17

Pearson's r correlations between estimates of skill mastery probabilities and total raw score

	VOC	EXP	IMP	INF	DIS	PRG	Total raw score
Using vocabulary (VOC)	1.00						
Using directly stated information (EXP)	.78	1.00					
Using implicitly stated information (IMP)	.70	.69	1.00				
Making connections (INF)	.73	.69	.80	1.00			
Separating ideas (DIS)	.82	.80	.67	.69	1.00		
Using culture (PRG)	.84	.86	.69	.72	.84	1.00	
Total raw score	.85	.87	.78	.79	.87	.89	1.00

All correlations significant at $p < .001$, $N = 102$

Additionally, on each skill a substantial proportion – between 20% and 60% – of participants received estimated mastery probabilities of .05 or less. Descriptively comparing raw total test scores with estimates of probability of mastery indicated that a useful raw score boundary for test profiles was 10 out of 38 (26%), below which probability of mastery estimates were consistently below .05. A second boundary was at a raw total test score of 21 (55%), above which participants consistently received varied mastery profiles which some skills showing a high probability of mastery and others a lower probability of mastery. Between these two boundaries, participants had some variation in skill mastery profile, but the probability of mastery for all skills was generally below .50.

3.7.1.2.2 Self-assessment tool

The mean scale scores among participants for each of the reported English reading skills ranged from .49 (self-assessed ability to make connections through inferencing, SA-INF, $SD = .18$) to .64 (self-assessed ability to understand discourse context and social situation, SA-PRG, $SD = .19$). The self-assessment scores on each skill were relatively normally distributed, with a somewhat negative skew and strong kurtosis. A full set of descriptives is given in Table 18.

Table 18

Descriptive statistics for self-assessment scores, by skill

Skill	Abbreviation	Mean	SD	Median	Range	N
Using vocabulary	SA-VOC	.60	.18	.58	.17 to .99	102
Using directly stated information	SA-EXP	.60	.19	.58	.08 to 1.00	102
Using indirectly stated information	SA-IMP	.55	.17	.50	.00 to .92	102
Making connections	SA-INF	.49	.18	.50	.08 to 1.00	102
Separating ideas	SA-DIS	.63	.20	.67	.08 to 1.00	102
Using culture	SA-PRG	.64	.19	.67	.17 to 1.00	102

Min = 0; Max = 1

In subsequent analyses, a single assessment score was used based on the raw scores. This choice was made because the correlations between items indicated moderate-to-strong correlations between all the items (see Table 19) with a coefficient alpha of .92, indicating a unidimensional structure. Therefore, the mean of all 18 items together was calculated for use in analyses. The distribution of the overall self-assessment scores was fairly normally distributed, with a mean of 3.37 ($SD = 0.63$) on a scale of 1 to 5.

Table 19

Spearman's rho correlations between self-assessment items

	voc1	voc2	voc3	exp1	exp2	exp3	imp1	imp2	imp3	infl	inf2	inf3	dis1	dis2	dis3	prg1	prg2	prg3
voc1	1.00																	
voc2	**31	1.00																
voc3	**31	***59	1.00															
exp1	***33	***51	***64	1.00														
exp2	**28	***52	***59	***67	1.00													
exp3	***40	***45	***37	***44	***40	1.00												
imp1	***36	**30	***41	***53	***52	***40	1.00											
imp2	***46	***40	***38	***34	***44	***33	*22	1.00										
imp3	***40	***35	***40	***49	***47	**26	***45	***42	1.00									
infl	**29	*25	**29	***35	**33	.09	**27	**32	***41	1.00								
inf2	***41	***43	***43	***47	***49	***38	***38	***48	*26	**29	1.00							
inf3	***36	**32	***41	***39	***42	*24	***36	***50	***35	***47	***47	1.00						
dis1	***34	***48	***50	***71	***60	***43	***50	**32	***45	***33	***54	***37	1.00					
dis2	***35	*26	***44	***44	***41	***42	***45	**26	***38	**31	***38	***40	***47	1.00				
dis3	***38	***43	***50	***67	***53	***48	***42	***36	***41	*22	***42	***48	***54	***53	1.00			
prg1	***41	***47	***58	***58	***53	***40	***37	***43	**31	.20	***56	***47	***64	***36	***59	1.00		
prg2	**30	**30	***53	***56	***51	***37	***43	***42	***38	***36	**32	***47	***53	***43	***47	***60	1.00	
prg3	***49	*21	***35	***30	*20	.10	***41	*21	**31	***37	**30	***35	**27	***34	***34	**29	*25	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$; $N = 98$ to 101

3.7.1.3 *Third meeting: Receiving the feedback report*

3.7.1.3.1 Planning section of report

The planning section of the report required participants to set goals, action plans, and monitoring plans. Participants' plans were coded in order to facilitate analysis according to the research themes.

For goal setting, goals were coded based on whether they were skill-specific, and whether they were observable. Being skill-specific meant that it was clear from the goal which reading skill it was addressing. This category was chosen because goal setting tends to be clearer when focused on a specific topic (Hattie, 2008). Observable goals were defined as those goals that were concrete enough that a learner could focus on and monitor progress toward them. This category was chosen because research also shows that observable goals facilitate self-regulated learning and positive learning outcomes (Nicol & Macfarlane-Dick, 2006).

For setting action plans, themes were coded for the action plans. Themes were developed based on those that emerged, and not every action plan was related to observed themes. Themes observed were study activities, using language from the report, referencing one's own life, and repeating action plans across multiple chosen skills.

In addition, specificity of action plans was coded based on how the specificity of plans grouped. Three main categories were identified, each of which was split into two or three subcategories. The first category consisted of responses with limited evidence of planning, and included 'plans' that were not actually plans, plans that indicated a general intention to study, and plans to practice a modality such as speaking or reading.

The second category included responses with moderate evidence of developing action plans. This category consisted of action plans that specified a general activity to practice a modality. It also included the action plans that identified a specific activity to practice a modality. Finally, the third category indicated specific activity plans. This category included the action plans that were general activities. It also included the action plans that were specific activities.

A variable for plan specificity was then developed based on the specificity categories previously identified, with non-plans scoring 0, general intention to study scoring 1, general intention to practice a modality scoring 2, plans to do language activities scoring 3, and specific, focused

activities scoring 4. As participants could write up to three plans, the specificity variable was the sum of each plan's specificity, with scores between 0 and 12.

Finally, for each skill that participants selected to work on, they were invited to write how they would monitor their progress. These monitoring plans were then coded for general themes, several of which arose: repeating plans across multiple skills, monitoring plans that were not actually monitoring plans but were goals or learning activities, and monitoring plans that used online activities. In addition, a major theme that arose was participants' locus of control. Some monitoring plans were self-regulated, while others had an external focus consisting of externally-regulated monitoring of performance: teacher, exams, exercises, and judgment by proficient speakers such as friends and family.

3.7.1.3.2 Eye-tracking and stimulated recall interview

As described in the data collection activities, 15 participants received their report while an eye-tracking device monitored where they looked. Immediately after these participants finished reading their report they also completed a prompted recall interview, in which the report and eye-tracking traces were used to prompt recollections of what they were thinking and feeling while they were reading.

The transcripts of the interviews were transcribed and a thematic analysis was conducted. Before describing the thematic analysis, it should be noted that of the fifteen participants, four participants were excluded from the thematic analysis. Three had very low English language proficiency, and their reported thoughts were entirely focused on attempting to understand the report content, for example: "Ok, understand description. That means, if you understand, don't understand it cr...what does description?...How relationship between the writer and reader? Relationship means how you between your relationship, yeah?" (participant with raw total score of 6 out of 38). Although other participants also paid attention to comprehension, these three participants were so focused on comprehension that little other cognition was reported.

In addition, one participant with low-but-adequate English proficiency misunderstood the task and thought the report was a timed task. As a result, she also reported little cognition other than seeking comprehension, and as her English proficiency was adequate, reported this comprehension-focused experience most clearly. For example [area of attention]:

[Introduction] No, nothing!... Yeah, I didn't think anything.... [EXP skill] I think I don't have any time to think other things... [IMP skill] I think maybe I didn't think anything... [DIS skill] I think I didn't think anything too! I just try to, I want to know the six skill, what is the six skill I can um, I can use... I read here and what do you agree, I just think yeah. What do you want to work on? What will help you? Yeah. That's very short, it's good. And find the main points, very important. But not think, I didn't think too much!... This sentence here I just maybe have, I think, oh that's very good. I can't think too much because I just want to go through the sentence. (EL, misunderstood task, raw total score of 10 out of 38)

Due to these four participants' intense focus on comprehension, themes concerning attention, cognition and affect that were observed among other participants were unobservable for this subgroup. In particular, all four participants appeared unmoved by their results, in contrast to the majority of the other participants. This equanimity may be because they knew their English proficiency was well below what was tested. However, it may also be because the cognitive load of comprehension was so great that other cognition and emotion was pushed out of view.

The transcripts of the remaining 11 stimulated recall interview participants were then coded for themes arising concerning attention, cognition and affect. There was a methodological challenge in coding, common in interviews, that some participants were verbose and provided long turns, while others were relatively taciturn. Additionally, some participants' transcripts contained more turns because these participants required more prompting and support than others during the interviews to think aloud and provide reported thoughts. These phenomena were particularly exacerbated due to the varying English language proficiencies of participants. Although I was able to talk in Mandarin Chinese with the Mandarin-speaking participants to ameliorate this issue, English was the only language I shared with the majority of other interview participants. Thus, simply counting the amount of speech and/or turns was not methodologically appropriate in this study, and therefore final variables developed used nominal categories.

All coding is guided by theory simply because the act of generalising requires drawing in the theoriser's pre-existing knowledge (Fink, 2000). Therefore, in some instances it is useful to start with an acknowledged theoretical basis to limit the influence of the researcher's prejudices (Fink, 2000; Hsieh & Shannon, 2005). This study had access to relevant existing theory through the

literature review used to develop the online post-report survey, so a somewhat directed approach to coding stimulated recall interview transcripts was adopted. As a starting point to guide coding, transcripts were roughly coded to identify whether similar constructs were observed to those hypothesized as processing outcomes on the survey: desire to use the report, feeling in need of help, trust in report content, reflecting on one's English skills, and sense of overwhelmedness (see Table 20 and Table 21). As noted, this choice was made because the theoretical decisions for the survey were based on what the literature indicated desired feedback report processing outcomes were, although differences between processes and outcomes were expected. Indeed, it was clear that there were some differences, for example feelings of overwhelmedness were not verbalized while reading the report, and many of the reported thoughts were focused on comprehending the report.

Table 20

Evolution of stimulated recall interview codes: initial tentative codes

Desire to use the report	Feeling in need of help	Trust in report content	Reflecting on one's English skills	Sense of overwhelmedness
--------------------------	-------------------------	-------------------------	------------------------------------	--------------------------

Table 21

Example coding using excerpts from two participants: initial tentative codes

Code	Participant TK	Participant UB
Desire to use the report	If this report helps me how to improve my skills and my progress, what I can do, what I will do. What I plan, like, to improve my, to see my progress... if I want to improve my progress, my skills, what I have to do, like, plan all my skills.	I was thinking, actually, to be honest, I was thinking, this one is useless because I will not just spend my time to guess and to see if I'm right. I will just read them all and guess some vocabularies and try to understand, but not- because this not useless, just I am too lazy to do it. Mhm.

Code	Participant TK	Participant UB
Feeling in need of help	I can tell the book, if after I read a book I can tell the story for myself and I can tell for others. And I enjoy the book and I go so fast, I finish the book. If it's harder and I don't want to refer often dictionary. But if I understand the first page without dictionary, I go so fast. Yeah, otherwise it's hard.	I was thinking, what's the word mean, because I didn't see that my opinion and the test opinion be compared these two part, so I was thinking what does it mean, do I agree or not.
Trust in report content	I was tired on that day because of test at home, I know I didn't do well. But I know. [How do you know you didn't do well?] Because I know myself... After reading everything, you understand yourself.	I was happy! Because the test give me a really confident. Because I'm not really confident about my total English levels so when I saw this I don't know if this is 100% or something, I'm just very happy so I look at this and I compare this two.
Reflecting on one's English skills	The first one, understand information that is suggested but not written out. It is a big up from my understanding. I understand in somehow. Instruction and description and the information, but here according to the test it say a big gap. Yeah... in this area I think big gap from my understanding.	Because I read this all very fast and when I read this I kind of forget what's the title so I come back to the title. And read them again!
Sense of overwhelmedness	No examples	No examples

As a result, the proposed themes were refined to: dealing with comprehension challenges, expressions of affect, reading to understand, and critical engagement with report content. The transcripts were then coded in detail for instances of these themes. Once coding had taken place, all instances of each code were viewed together, and considered in the light of literature on self-regulation. Two very broad themes noted consistently in the literature appeared to delineate coded speech: affect and cognitive strategies. Codes in the themes 'dealing with comprehension challenges' fell into both, while 'expressions of affect' codes were clearly affective. 'Reading to understand' and 'critical engagement' codes were clearly cognitive. This round of coding is summarized and exemplified in Table 22.

Table 22

Evolution of stimulated recall interview codes: intermediate revised codes

	Dealing with comprehension challenges	Expressions of affect	Reading to understand	Critical engagement with report content
Affective	X	X		
Cognitive	X		X	X

Another round of coding took place to further clarify themes. Affective strategies emerged as whether participants minimized negative results, and whether participants reported emotional responses. Cognitive strategies that emerged were the extent to which participants related the report to their own lives, how participants dealt with comprehension issues, and the type of critical evaluation of the report that participants engaged in. The codes that emerged, together with their categories, are summarized and exemplified in Table 23 and Table 24.

Table 23

Evolution of stimulated recall interview codes: final codes with categories

Affective		Cognitive		
Minimized poor results	Emotional	Related report to own life	Dealt with comprehension challenges	Critically evaluated report content
Yes / No	Yes / No	No / Some / Lots	None reported / Wanted help / Self-resolved	No / Yes then content accepted / Yes then accepted some content, rejected some

Table 24

Example coding using excerpts from two participants: final codes with categories

Code	Participant TK	Participant UB
Minimized poor results	I was tired on that day because of test at home, I know I didn't do well. But I know. [How do you know you didn't do well?] Because I know myself... After reading everything, you understand yourself.	
Emotional		I was happy! Because the test give me a really confident. Because I'm not really confident about my total English levels so when I saw this I don't know if this is 100% or something, I'm just very happy so I look at this and I compare this two.
Related report to own life	I can tell the book, if after I read a book I can tell the story for myself and I can tell for others. And I enjoy the book and I go so fast, I finish the book. If it's harder and I don't want to refer often dictionary. But if I understand the first page without dictionary, I go so fast. Yeah, otherwise it's hard.	
Dealt with comprehension challenges		Because I read this all very fast and when I read this I kind of forget what's the title so I come back to the title. And read them again! I was thinking, what's the word mean, because I didn't see that my opinion and the test opinion be compared these two part, so I was thinking what does it mean, <i>do I agree or not.</i>

Code	Participant TK	Participant UB
Critically evaluated report content (or did not do so)	<p>If this report helps me how to improve my skills and my progress, what I can do, what I will do.</p> <p>What I plan, like, to improve my, to see my progress... if I want to improve my progress, my skills, what I have to do, like, plan all my skills.</p>	<p>I was thinking, actually, to be honest, I was thinking, this one is useless because I will not just spend my time to guess and to see if I'm right. I will just read them all and guess some vocabularies and try to understand, but not- because this not useless, just I am too lazy to do it. Mhm.</p>
	<p>The first one, understand information that is suggested but not written out. It is a big up from my understanding. I understand in somehow. Instruction and description and the information, but here according to the test it say a big gap. Yeah... in this area I think big gap from my understanding.</p>	

As counting was not a reliable means of interpreting the strength of each of the five themes for the participants due to individual differences in verbosity, themes were mostly divided into nominal categories. The exception was 'relating report to one's own life', which emerged into a three-category ordinal organization of no/some/lots. Some participants did not connect the report to their own lives, and were categorized 'no'. Participants who made only tenuous or one-off connections were categorized as 'some', while those who consistently referred to their own lives were categorized as 'lots'. As for the other two cognitive strategies observed, participants were consistently observed to deal with comprehension issues in two different ways: express lack of understanding and desire for assistance, or resolve comprehension challenges themselves. A third category consisted of those participants who reported no comprehension challenges. Finally, three types of critical evaluation of report content were observed: no evidence of critical evaluation, questioning content but eventually accepting it, and questioning content then accepting some and rejecting some. Regarding the two affective strategies, both were binary; the participants either reported the strategy or they did not.

Table 25 shows how interview participants reported interacting with the report according to the identified themes.

Table 25

Participant interaction profiles: stimulated recall interviews

Participant	Minimized poor results	Emotional	Related report to own life	Dealt with comprehension challenges	Critically evaluated content
1	Yes	No	Some	Wanted help	Yes, but accepted all
2	No	Yes	Lots	Self-resolved	Yes, both accepted and rejected
3	Yes	No	Lots	Wanted help	Yes, but accepted all
4	No	No	Some	Self-resolved	Yes, but accepted all
5	Yes	Yes	Some	Wanted help	No
6	Yes	Yes	No	Self-resolved	Yes, both accepted and rejected
7	No	Yes	No	Wanted help	No
8	No	Yes	No	None reported	Yes, both accepted and rejected
9	Yes	No	No	Wanted help	Yes, both accepted and rejected
10	Yes	No	Lots	None reported	No
11	No	No	Some	Wanted help	No

The ways in which these observed themes were analysed and synthesized to address the research themes are described in the section that describes data analyses by research theme.

3.7.1.3.3 Post-report survey

Like the background survey psychology items and the assessment items, several groups of items on the post-report survey needed consolidating into composite variables in order to be useful for subsequent analyses. Other items required preparation prior to use in analysis. Validation, consolidation and preparation of the variables are described below.

3.7.1.3.3.1 Attention

One area of interest was to gauge the comprehensiveness of participants' attention to the report. To do this, binary variables were first created to identify the extent to which participants reported a lot of attention on each section of the reports. In the binary variables, responses of 4 or 5 ('A lot of time', 'All my time') were considered 'a lot of attention' (1) and responses of 1 to 3 ('I did not look at this', 'A little time' and 'Some time') were considered 'little attention' (0). Based on these binary variables, a proxy variable for comprehensiveness of attention was then created that described how many report sections participants had paid 'a lot of' attention to.

3.7.1.3.3.2 Cognition and affect with respect to amount of information provided

For the eight items intended to explore participants' cognition and affect with respect to the amount of information provided on the report, Spearman's *rho* correlations were used to identify whether the items formed identifiable groups. Note that the raw items were considered ordinal scales given their 1-5 rating scale and therefore correlated using Spearman's *rho*. As with the goal orientation items, an exploratory factor analysis was also conducted (see Appendix 6), but as previously the study sample size ($N = 102$) was too small for a robust exploratory factor analysis, so final composite variables were not based on factor scores derived from these results. However, the exploratory factor analysis results were similar to the factor structure observed in the simple correlation matrix, and therefore the results support the use of composite variables based on mean scores.

The correlations indicated one adequately cohesive group of three items: sense of overwhelmedness. The items in this group were 'There is too much information to understand' (item A1), 'There is so much information I cannot remember everything' (item A2), and 'I only thought about the very interesting parts' (item B4) ($\rho = .21$ to $.37$, $p < .05$). Note that coefficient alpha was $.55$. No other group of three or more items correlated, so other items were dropped. Dropped items were 'I looked at and thought about all parts of the report' (item B2), 'I looked at the different parts of the report and thought about how they connected' (item B3), 'Some parts of the report were more interesting than others' (item B1), 'The information is very general and I want more details' (item A4), and 'I was looking for more information in the report' (item A3). The result was that no variable for the construct of comprehensive reflection could be developed. The full set of correlations is shown in Table 26.

Table 26

Spearman's rho correlations among items assessing cognition and affect for amount of information

	A1	A2	B4	B2	B3	B1	A4	A3
A1	1.00							
A2	** .27	1.00						
B4	* .21	*** .37	1.00					
B2	-.18	-.04	*-.23	1.00				
B3	.18	.13	.06	*** .39	1.00			
B1	-.11	.19	*** .34	.14	.20	1.00		
A4	-.09	.15	-.10	* .21	.15	-.02	1.00	
A3	.07	*** .34	* .25	.15	* .23	.17	*** .37	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$; $N = 92$ to 100

3.7.1.3.3.3 Cognition and affect with respect to report content

For the eight items intended to explore participants' cognition and affect with respect to the content of the report, Spearman's *rho* correlations were also used to identify whether the items formed identifiable groups. As with the items addressing attention, note that the raw items were considered ordinal scales (1-5 rating scale) and therefore correlated using Spearman's *rho*. An exploratory factor analysis was also conducted (see Appendix 6), but once again study sample size ($N = 102$) prevented a robust exploratory factor analysis, so final composite variables were not based on factor scores derived from these results. However, the exploratory factor analysis results were similar to the factor structure observed in the simple correlation matrix, so the results support the use of composite variables based on mean scores.

The correlations indicated two cohesive groups of three items each: trust in the report content and reflecting on one's English skills. The item group for 'trust in the report content' consisted of 'The test results are more correct than my self-assessment' (item A2), 'I believe what the report says about my English skills' (item A3), and a reverse-coded 'The report is wrong about my English skills' (item A4) ($\rho = .29$ to $.53$, $p < .01$). Coefficient alpha for this group was $.61$. The item group for 'reflecting on one's English skills as a result of reading the report' consisted of 'I am thinking about my English skills now' (item B2), 'I am thinking about how I can use the report for my English studies' (item B3), and 'I am thinking about how I am studying English

now' (item B4) ($\rho = .38$ to $.57$, $p < .001$). Coefficient alpha for this group was $.70$. Two items were omitted because they correlated with all items: 'I will do what the report told me to do' (item A1) and 'I am comparing the report with my ideas about my English skills now' (item B1). The full set of correlations is shown in Table 27.

Table 27

Spearman's rho correlations among items assessing cognition and affect for report content

	A2	A3	A4(r)	B2	B3	B4	A1	B1
A2	1.00							
A3	***.51	1.00						
A4(r)	** .29	***.53	1.00					
B2	** .32	***.34	.17	1.00				
B3	*.22	*.24	.19	***.38	1.00			
B4	*.24	** .27	.17	***.49	***.57	1.00		
A1	***.39	***.48	** .26	** .30	***.33	***.48	1.00	
B1	.13	** .30	*.25	*.25	** .27	** .31	***.49	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$; $N = 94$ to 101

3.7.1.3.3.4 Planned usage

For the eight items intended to explore participants' planned usage of the report, as with other processing item groups, Spearman's *rho* correlations were used to identify whether the items (assumed to be ordinal on a scale of 1 to 5) formed identifiable groups. As with other item groups, an exploratory factor analysis was also conducted (see Appendix 6) but final composite variables were not based on factor scores derived from these results due to the small sample size of the study ($N = 102$). However, as with previous exploratory factor analyses, the results were similar to the factor structure observed in the simple correlation matrix, supporting the use of composite variables based on mean scores.

The correlations indicated two cohesive groups of three items each: desire to use the report and feeling in need of help to use the report. The item group for 'desire to use the report' consisted of 'I want to use the report feedback in my English studies' (item B1), 'I hope to use my report to change my English study activities for better learning' (item B2), and 'I want to use the report's suggestions in my English studied' (item B3) ($\rho = .53$ to $.61$, $p < .001$). Coefficient alpha for this group was $.72$. The item group for 'feeling in need of help to use the report' consisted of 'I need help to start using the information in my report' (item A2), 'I need someone to tell me how to

move forward with my English studies now' (item A3), and 'I want to use my report for my English learning but I'm not sure how' (item A4) ($\rho = .22$ to $.61$, $p < .05$). Coefficient alpha for this group was $.74$. Two items were omitted because they failed to consistently correlate with any group of items: 'I know how to use the report for my English learning' (item A1), and 'I will continue with my English studies without using this report' (item B4). The full set of correlations is shown in Table 28.

Table 28

Spearman's rho correlations among intended use items

	A2	A3	A4	B1	B2	B3	A1	B4
A2	1.00							
A3	***.61	1.00						
A4	***.44	*.22	1.00					
B1	***.41	** .28	.11	1.00				
B2	** .30	***.47	.10	***.61	1.00			
B3	*.26	***.33	.07	***.55	***.53	1.00		
A1	.02	.12	-.02	*.26	***.34	***.34	1.00	
B4	.03	-.08	** .31	-.06	-.08	** -.32	.00	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$; $N = 94$ to 100

For the seven binary items that captured what participants' intended to discuss with others about the report, item correlations were viewed using phi coefficients, due to the binary values of the variables. The topics did not consistently cluster in terms of correlation (shown in Table 29), so no composite variables were developed.

Table 29

Phi coefficient correlations among discussion topics (clustered where possible)

	S1	S2	S7	S5	S3	S4	S6
S1	1.00						
S2	** .29	1.00					
S7	***.36	*.21	1.00				
S5	*.24	.11	** .27	1.00			
S3	.18	-.07	.16	***.46	1.00		
S4	.11	.15	***.33	.19	*.21	1.00	
S6	.08	** .27	.16	.09	-.04	** .26	1.00

* $p < .05$; ** $p < .01$; *** $p < .001$; $N = 100$

3.7.1.4 *Fourth meeting: Delayed recall interview*

In total, 15 participants completed a delayed recall interview, and their interviews were then transcribed. For analysis, one participant's responses were not included because he had returned to his originally country shortly after receiving the report and returned only recently; during the past month he had done nothing related to English. The responses of two other participants were also not reported because their test scores were very low (6 and 8 out of 38 items correct), indicating lower proficiency than was intended for the study, and additionally they were elderly and struggling with memory. As a result, these two participants recalled nothing of the reports when asked. For the remaining 12 participants, to find out which aspects of the report were sufficiently paid attention to that they were recalled without prompting one month later, the transcripts were coded for references to report sections. The ways in which these references were analyzed to address the research themes are described in the section describing data analyses by research theme.

To find out how the 12 participants reported using their reports during the month since they received them, participants' transcripts were first viewed to identify who used the reports and who did not. Next the responses of those who did not use the reports were searched to identify why they did not use them. Third, any consistent comments or themes arising among the participants were noted. Finally, ways in which the participants reported using their reports was extracted from the transcripts. It should be noted that the delayed recall interviews were also found to be a source of information about participant motivation and daily environment, although they were not specifically designed to elicit this information. However due to this information, the interviews were additionally useful for elaborating understanding about factors underlying participant usage of feedback. Analyses based on this data are described in the section describing data analyses by research theme.

3.7.2 **Data analyses by research theme**

As previously indicated , quantitative and qualitative analyses were conducted to provide data that informed each of the research themes. The analyses used to address each theme are now discussed in further detail.

3.7.2.1 *Theme 1: Characteristics of adult immigrant English language learners in Canada*

Theme 1 explored the characteristics of adult immigrant English language learners in Canada. Background survey items relating to occupation, English usage, and aspirations for future English proficiency were analysed descriptively. In addition, comments about these topics were provided spontaneously by some participants during eye-tracking and recall interviews. Although participants were not specifically asked about these topics, themes were noted from the comments that were offered.

The composite variables for goal orientation were analysed descriptively to identify overall trends among the participants in terms of goal orientation. To identify typical goal orientation profiles for the study participants, a latent profile analysis (Lazarsfeld & Henry, 1968) was conducted. Models using the three goal orientation variables were developed, with different numbers of classes used in different models. The best fitting model was identified using the lowest adjusted Bayesian information criterion (BIC) and the highest loglikelihood value (smallest chi-square value). In addition, beliefs about intelligence were analysed descriptively. Comments from interview participants who held strong beliefs about intelligence or goal strong goal orientations, that represented these beliefs and orientations, were noted.

Next, the probabilities of mastery for each reported skill, the raw total test scores, and the self-assessment scores were all analysed descriptively. In addition, using the probabilities of mastery for each of the six identified English reading skills, participants' typical English reading proficiency profiles were identified through latent profile analysis. Models were developed using the probabilities of mastery for each of the six skills identified by the cognitive diagnostic model: using vocabulary, using explicitly stated information, using implicitly stated information, inferencing, distinguishing ideas, and using cultural knowledge. Note that due to an administrative error, some participants initially received probabilities of mastery that were somewhat lower than their true estimated probabilities of mastery. However, the latent profile analysis used their correct probabilities of mastery in order to identify typical English reading proficiency profiles.

Several models were developed using the six skills, with different numbers of classes used in each model. The best fitting model was identified using the lowest adjusted BIC and the highest loglikelihood value (smallest chi-square value). However, models with six classes and over did

not converge, likely due to insufficient sample size. As a result, it is possible that more classes were observable, but were not identifiable in this relatively small dataset.

Next, participants' self-assessment scores were analysed descriptively together with participants' raw total test results and placed into a latent profile analysis to identify whether groups of participants emerged who substantially over- or underestimated their English reading proficiency. As with previous analyses, several models were developed, all of which used two variables (total raw score and overall self-assessment score) but the number of hypothesized classes changed in each model. The best fitting model was identified using the lowest adjusted BIC and the highest loglikelihood value (smallest chi-square value).

Finally, prior research has often linked language learners' social environment, psychological factors, and language proficiency in various ways. Therefore, in order to open a window into the network of relationships that existed among the individual characteristics observed in this study, a correlational analysis was conducted for those characteristics observed in the form of quantitative variables. Goal orientation and beliefs about intelligence, total raw test score, self-assessment score, activities using English, confidence about English learning, and gender, year of birth and level of education were all observable in this manner.

3.7.2.2 Theme 2: Relationship between attention to and processing of feedback

Theme 2 explored the relationship between attention to and processing of feedback. Fixations recorded for each participant during eye tracking were analysed descriptively. As there was much variation among participants regarding how much time they had spent looking at the report, the percentage of time spent looking at each section was calculated, in order to identify which sections received relatively more attention. Eye-tracking paths were also viewed in order to identify similarities and variations in how participants read through their reports.

As reported earlier, fifteen participants completed the delayed recall interviews. After identifying which parts of the report were recalled by each participant, the themes among those who recalled each section were explored. For example, the suggestions recalled were ones that participants found personally meaningful, and specificity of plans also varied. Based on frequency with which each section was recalled, a hierarchy of recall was determined.

3.7.2.3 *Theme 3: Language learners' affective and cognitive interaction with report content*

Theme 3 explored language learners' affective and cognitive interaction with report content. The categories for affective and cognitive processing strategies that were derived from coding were described and explored for internal differences. In addition to observing in detail how participants interacted with their reports, the outcomes of these interactions and any subsequent cognitive processing were of interest because the thoughts and feelings that language learners take away from their report will affect how they use their feedback.

Therefore, to investigate how the processing outcomes observed on the survey were related, a latent profile analysis was conducted to identify the ways in which participant profiles clustered. Models were developed iteratively in logical steps. First, several theoretically plausible sets of variables were identified. The best fitting number of classes was then identified for each set of variables using the lowest adjusted BIC among each set of models. Finally, these class best-fit models were compared with each other using the highest loglikelihood value (smallest chi-square value).

3.7.2.4 *Theme 4 – Usage of feedback*

Theme 4 explored participants' usage of feedback. In terms of usage of the planning section of the report, the number of skills chosen was analysed descriptively. How participants set goals, action plans, and monitoring plans was then also analysed descriptively, including whether behaviour varied across skills, and across skill priority (that is, whether it was the first, second or third skill selected). Planned discussion topics and intended discussion partners were analyzed descriptively. Wilcoxon-Mann-Whitney tests were then conducted to see whether participants selecting a specific topic were more likely to want to talk to certain people.

After one month, participants who were still willing and available, met and talked about what they recalled of the report, and whether they had used it. Among the recollections and reported usage for the 12 participants whose interviews were usable, several themes were identified as discussed earlier. For example, one theme was that the report had said participants' English reading skills were lower than they expected, another theme was rejection of the report, while another was lack of time to use it. These themes were described, and then participants' reported activities were compared to what they said they would do in the planning section of their report.

A major question for report usage is what factors impact usage. To answer this question, inferential statistical analyses were conducted, including correlational analyses, a Kruskal-Wallis test, and linear, ordinal and logistic regression analyses. Note that given the diversity observed within the participant population, and the ordinal and/or highly skewed nature of many of the variables used in the analyses, non-parametric tests were used for most analyses. In one instance a measure was considered continuous and the assumption of normality was not substantially violated, so a linear regression analysis was used due to the ease of interpretability of linear regression results. Overall, the analyses determined the relationship of observed psychological characteristics, reported attention and reported processing experience with usage of the report planning section, and with intended discussion interlocutors and topics.

3.8 Conclusion

This description of my research methods introduced the context and participants, and described the research design and data collection procedures. It reported on the instruments used to collect data, how composite variables were developed for quantitative data sources, and how coding was implemented for qualitative data. Finally, analyses were conducted to address the research themes described. In the next chapter, the results of these analyses are reported.

Chapter 4: Results

4.1 Theme 1 – Individual characteristics

In the first research question, the background characteristics that individual language learners bring to their interaction with the report are investigated. All language learners bring a variety of social and psychological backgrounds to any learning experience. This is especially true for a heterogeneous group such as adult immigrant English language learners who are receiving a report from a test that is not embedded in a particular program of instruction, as was the case in this study. It is possible that these background characteristics underlie many of the variations in processing experience that the study participants reported and influenced how learners subsequently used their reports.

For this reason, the social and psychological backgrounds of the adult immigrant English language learners participating in this study are now considered, in order to understand how these factors are consistent or vary among the study population, and to gain an understanding of how these factors may be related. This information will be valuable for use in explaining why adult immigrant learners used the reports in different ways.

4.1.1 RQ 1.1: What are the occupational contexts of adult immigrant English language learners in Canada?

First, the participants in this study were all English language learners in southern Ontario who had either immigrated to Canada or who were currently living in Canada and planning to remain in Canada in future. They had all arrived within the past year, and the majority (78%) were currently in government-funded English study programs. As such the study participants brought a distinctive set of characteristics that are commonly found among immigrant language learners. Moreover, Canadian immigration policy, in a quest to boost the country's population, is distinctive for proactively soliciting large numbers of highly educated immigrants and their dependents as well as relatively large numbers of refugees. As a result, this study's participants represent a distinctly Canadian immigrant population. However, within this distinct population, there was wide variation in past and present occupations and daily activities.

As would be expected given Canada's wide ranging but professionally-focused immigration policy, participants' work experience prior to coming to Canada was diverse but predominantly

white collar; 20% of participants reported teaching and education as their occupation prior to coming to Canada, and a further third (34%) reported administration, health or social services, or scientific/technical areas as their background. Notably, 15% reported being homemakers or retired prior to immigration; these participants reflect the family reunification policy aspect of Canadian immigration policy.

Note, however, that not all the participants expected to continue their previous occupations in Canada. Anecdotally, a frequent question when filling out the background survey was “Do I write my occupation before Canada or what my occupation will be in Canada?” Although not directly questioned about occupation and career, some recall and eye-tracking interview participants mentioned plans for a change of occupation. Some participants were using immigration as an opportunity for a career change, for example: “I’m planning to take more courses specific in my area...And I want to, I want to get a job in the future...Or planning to open my own business...But now, I boring! I want to do something different in my life! [Not do the same job I did before.]” Others believed they would be unable to continue their previous career, for example:

Maggie: Do you plan to work in the same profession that you worked in Iran?

Participant: Maybe not. Because it’s very difficult here to find a job in my field. Maybe I change it.

Maggie: Oh. What was your field?

Participant: Sales and marketing.

Maggie: You have to have really excellent English.

Participant: Yes! Because of this I have to change!

In terms of current daily occupations in Canada, reported activities reflected the recently arrived status of participants, the fact that three quarters were studying English almost full-time, the adult age group (median age 38) and the female-dominated gender distribution (74%) of participants. Most participants (84%) had recently been studying English, and half (51%) reported caring for family. A smaller proportion were working; 10% of the participants reported

working in a store or restaurant, 9% reported working in an office, 3% reported working in a factory and 2% reported working in construction.

4.1.2 RQ 1.2: What are the English language environments of adult English immigrant language learners in Canada?

All the participants were living in metropolitan areas of southern Ontario in which English is the common language of communication outside of immigrant-language communities. Therefore, English language usage was generally frequent, but the extent of usage varied among participants.

When asked about their English usage, a majority of participants reported shopping (78%), watching TV or video (76%), and reading news (70%) three or more times a week in English. These activities reflect a typical Canadian context: the use of English for commerce, the prevalence of television in Canadian households, and the accessibility of online news and free print newspapers.

English usage activities with moderate engagement in this population were reading books and using email, and talking with friends. Half of the participants reported reading books (53%) and/or writing emails (53%), and about two thirds (64%) reported talking with friends, three or more times a week in English. Such activities are likely less prevalent because there is no need to use English for these tasks or the tasks themselves are not done at all by some participants. For example, it is plausible to only have friends who speak one's own language, only use email to communicate socially (or not have email), and to not read books due to lack of time or interest. However, all of these activities were frequently referenced by interview participants as activities they already did or felt they should do to practice their English.

Finally, social media, chatting online, and talking with family were least likely to take place in English; only about one third of the participants reporting doing these activities regularly in English (30%, 29% and 29% of participants, respectively). Once again, engagement in these activities in any language was unlikely for some participants who had low digital literacy or were not living with their families. Others would be unlikely to use English for these generally informal and social – and therefore likely interacting with a known-language interlocutor – activities. Note also that the only writing activity included in the analysis was email; other writing activities were infrequently reported because other writing activities were situated in

work or educational contexts, and at least three quarters of the participants were not yet operating in these contexts.

An important environmental characteristic of this learner population is that participants used their social resources, particularly family where available, to assist in their language learning. Nine of the twelve delayed recall interview participants reported using these types of resources; of the three remaining participants, two were probably living alone. The mothers with school-age children and younger frequently utilized their children's schoolwork or learning for their own English acquisition. For example:

And sometimes [the teacher] give [my daughter] writing, you know English also. My husband used to say you can check, you can improve English. So I check for [my daughter], I usually read her story. How she write. It helps me also...

And cartoons with my son, yeah I've learned a lot of things. Because most of the time our TV is on and my son watch cartoons in English language and I learned a lot of words and sentence from his cartoon.

Participants also utilized their family as mediators when negotiating English communication tasks and as cheerleaders and supporters when negotiating difficult tasks:

Sometimes I talk to the other Canadian and we talk to three with my son. But I want to talk to him together so I try. Sometimes I am just watching but sometimes I try but I ask my son. But it's not weird, it's not bad? I ask him. Oh yeah, ok, you can do it, continue [he says]. He make encourage. So I try.

And I try to, I write the email or video with my husband who live now in the Chinese, I try to use English because it can help me to remember the word, yeah?... He can't speak. But his writing is better than me.

As for those participants living alone, it is likely they rely more heavily on resources outside the home, for example:

I have a neighbour, and when I saw her I try to speak with her. She is a Philippine but she live here for a long time so she speak very well English and I practice with her. I'm

practice with her my speaking, my writing too because I have her phone number, sometimes I wrote and chat with her.

It must also be noted that several participants mentioned personal circumstances that prevented them devoting more time to their English language studies, principally childcare and household management responsibilities, as well as specific events such as family visits or family problems. It is a reality for immigrants with family care duties that utilizing study opportunities can be challenging without appropriate financial and logistical support.

In summary, the participants in this study often reported similar social contexts, in that they were living in more-or-less English speaking communities, and utilized the resources available to them. However, they also differed by the types of English activities they engaged in regularly. There was also variety among participants in the other responsibilities they were dealing with in addition to English language studies.

The social and linguistic contexts of language learners exert powerful influence on how learners engage with their learning, and the extent to which they engage with their learning. In this respect, an additional contextual aspect is that of learners' motivations and aims for English learning – and their confidence in achieving those aims. This topic is addressed next.

4.1.3 RQ 1.3: What are the English language goals of adult immigrant English language learners in Canada?

Despite the variation in English usage, resources and context, expectations about English language achievement among study participants were consistently high, which is to be expected given the high stakes of mastering English in the participants' immigrant context. However, confidence in achieving these aims was more variable. While 95% of the participants aspired to watch TV or movies in English, 73% were confident they would do so. Likewise, 90% aspired to read novels in English, but 65% were confident they would be able to do this. Fully 97% of the participants aspired to make friends in English, while only 63% were confident they would do so. Confidence about attaining in English-based employment was even lower; 44% were confident they would achieve this, although 95% of the participants aspired to achieve employment.

Likewise, participants were highly motivated to learn English. As previously noted, many of the participants aspired to work, and were learning English either to get a job directly, or to enroll in

further training. For example: “I hope I find a good job. But first I have to improve my English, pass some course, and maybe I can find. I hope!” and:

I’m planning to get enrolled in the university. I’m not yet confirmed whether it’s the masters program or the doctoral program. I will decide it later based on the score of my IELTS or TOEFL exam...As long as I am, I am continuing with my special drive that I will not stop until I reach my target, so I think I’m doing pretty well with my continuous drive I think.

Others were simply motivated to fully participate in Canadian society and judged their progress in English based on the extent of successful participation, for example: “For example when I went to clinic, my husband didn’t come with me. I could do everything by myself. Yes, register and after wait, talking and conversation with doctor, I didn’t have any problem” and:

Because yesterday, before yesterday I went to my condo [corporation meeting]. ...And I went there, I talk with people, I understanding more, yeah. And my son said oh you good, and so yeah. But just one time I think so! I’m not yet more understanding exactly. I need, because the, who some guy, some people said? But I don’t understand that time. So I continue. Continue. I need to continue study.

Together with the social context of learners, these motivations and aspirations will impact learners’ expected progress, desired learning goals and willingness to persist. However, another major influence on learners’ engagement and learning outcomes is individual psychological factors. The factors observed in this particular study are now described.

4.1.4 RQ 1.4: What are typical goal orientation profiles and beliefs about intelligence among adult immigrant English language learners in Canada

The individual psychological factors considered in this study were goal orientation, beliefs about intelligence, and metacognitive control. For goal orientation, a theory which reflects underlying social motivations for why individuals learn, three constructs were used: mastery orientation, performance prove orientation, and performance avoid orientation (Button, Mathieu, & Zajac, 1996; Elliot & Harackiewicz, 1996; Midgley et al., 1998).

Learners with mastery orientations are learning in order to master the material. In this study the mean score (on a scale of 1 to 5) for a mastery orientation was 4.28 ($SD = 0.54$). Notably, the

minimum score was 2.86, indicating a strong negative skew in mastery orientation for this population. Examples of mastery-oriented thinking observed during interviews included: “I was thinking how to plan myself, um, maybe, this when I was reading this maybe um. For example, I have problems with idioms, so I can establish, I could establish a goal.” “Mechanize my logic in English is my main goal.”

Learners with performance prove orientations are seeking to prove to others that they are capable, and in this study the mean score for a performance prove orientation was 2.71 ($SD = 0.84$). An example of performance prove-oriented language was: “I mean the improvement, means in which level, the test was. What was the, the benchmark level.”

Learners with performance avoid orientations are primarily seeking to avoid looking stupid to others, and in this study the mean score for a performance avoid orientation was 2.45 ($SD = 0.81$). An example of a comment from a participant with a strong performance avoid orientation was: “But also it’s the first time I read this one. In [my teacher’s] class is different, no? The paper she gave, it’s a different test, that’s why, that’s the reason I think I thought.” Both performance prove and performance avoid were normally distributed, with a slight positive skew.

To identify typical goal orientation profiles for the study participants, a latent profile analysis was conducted using the three goal orientation variables developed. The best fitting model had five classes (adjusted BIC = 632, loglikelihood value = -298, $N = 101$) and an entropy of 0.84, indicating adequate classification quality. However, models with six classes and over did not converge, likely due to insufficient sample size. Therefore, it is possible that more classes were observable, but were not identifiable in this relatively small dataset. Model comparisons are presented in Table 30. Figure 6 shows, for each of the five classes in the best-fitting model, the mean score for each goal orientation.

Table 30

Participant goal orientations: latent class profile models comparison table

Number of classes	Adjusted BIC	Loglikelihood	Entropy	<i>N</i>
2	646	-314	0.80	101
3	641	-308	0.75	101
4	638	-304	0.81	101
5	632	-298	0.84	101
6 **				

** Possible non-identification of model. The chosen model is in bold type and dark grey highlight.

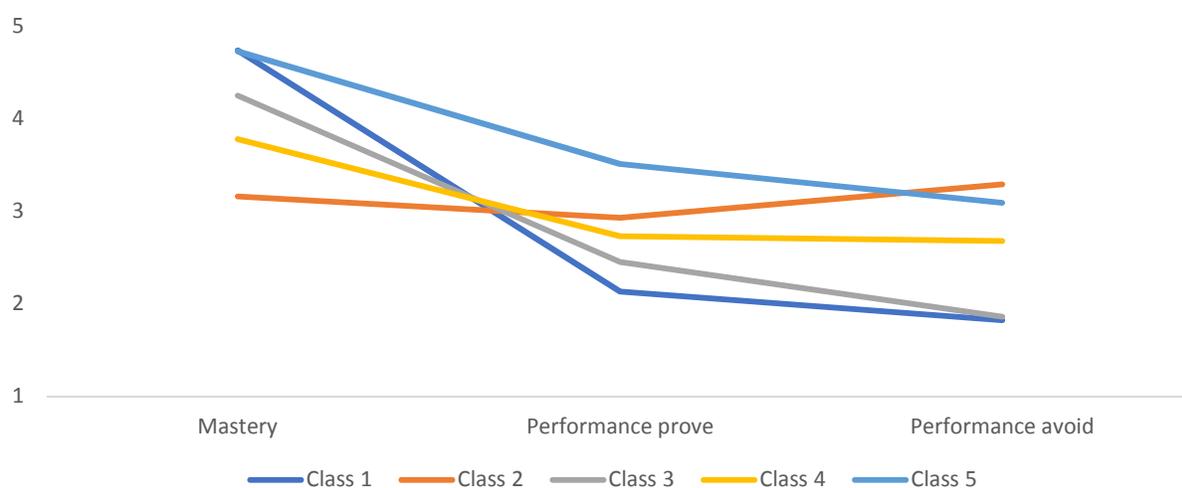


Figure 6. The mean scores on each goal orientation, for each class

Participants in Class 1 ($n = 26$) clearly held very strong mastery goal orientations and weak performance orientations, as the mean mastery score was 4.74 ($SD = 0.18$) on a scale of 1 to 5 compared to a mean performance prove score of 2.13 ($SD = 0.72$) and a mean performance avoid score of 1.82 ($SD = 0.50$), both also on a scale of 1 to 5.

A similar class was Class 3 ($n = 18$), whose participants in general also held quite strong mastery orientations and weak performance orientations. However, participants in Class 3 held on average weaker mastery orientations ($M = 4.25$, $SD = 0.17$) and higher performance prove orientations ($M = 2.45$, $SD = 0.64$) than participants in Class 1. The mean performance avoid orientation was, similar to Class 1, very low ($M = 1.86$, $SD = 0.52$).

The remaining three classes consisted of participants holding substantially higher performance avoid goal orientations. Class 4 ($n = 27$) continued the trend from Classes 1 and 3, with even lower mean mastery orientation ($M = 3.78$, $SD = 0.21$), and higher performance prove orientations ($M = 2.73$, $SD = 0.67$). A substantive difference between Class 4 and Classes 1 and 3 was that the mean performance avoid goal orientation for Class 4 ($M = 2.68$, $SD = 0.73$) was higher than for the other two classes.

Class 2 ($n = 7$) was unique in that the highest mean among the three orientations was for performance avoid ($M = 3.28$, $SD = 0.58$). Participants in Class 2 also held, on average, the lowest mastery orientations ($M = 3.16$, $SD = 0.25$). The mean performance prove score was moderately high ($M = 2.93$, $SD = 0.54$).

Finally, Class 5 ($n = 23$) was another unique class, in that the mean mastery orientation was very high ($M = 4.73$, $SD = 0.20$) – as high as Class 1, yet the performance orientation means were also high. In particular, the mean performance prove orientation was the highest of all the classes ($M = 3.51$, $SD = 0.75$), and the mean performance avoid orientation ($M = 3.09$, $SD = 0.56$) was second highest, and almost as high as Class 2.

In summary, performance prove and mastery goal orientations generally contrasted with each other among the classes identified; classes with higher mean performance prove orientation had lower mean mastery orientation. However, mean performance avoid orientation varied across the classes. In terms of specific classes, Classes 1 and 3 held very high mastery orientations and very low performance prove and avoid orientations, indicating a strong mastery orientation. In contrast, Classes 4 and 5 held performance prove orientations that were closer in strength to their mastery orientations, although Class 4 held much weaker orientations than Class 5 in general. Both classes held performance avoid orientations of similar strength to their performance prove orientations, indicating a general performance orientation. Finally, Class 2 held the weakest mean mastery orientation, on par with its mean performance prove orientation. Moreover, the strongest mean orientation for Class 2 was performance avoid, a unique characteristic among the classes, and indicative that participants in Class 2 held a dominant performance avoid orientation.

In addition to goal orientations, participants were asked about their beliefs about intelligence. Beliefs about intelligence fall on a spectrum ranging from believing that intelligence is innate and fixed at birth, to believing that intelligence is malleable and can change incrementally over

time. Although participants were asked several questions on the background survey about these beliefs, a single item stood out as best capturing participants' beliefs on the topic in the language learning context. This item asked participants how much they agreed with the following statement: 'People are either good or bad at learning languages.' The mean response on this item was 3.0 (on a scale of 1 to 5, $SD = 1.23$), with a range of responses on either side.

Two comments from a participant in her late forties and a participant in her late sixties demonstrated these mindsets in regards to language learning and age. Both of them struggled to remember new vocabulary. The participant in her late forties held strong fixed beliefs about intelligence, and noted: "Old is, usually forty over, remember is very bad." In contrast, the participant in her late sixties held strong incremental beliefs about intelligence, and these beliefs are reflected in her comment: "I think it is better to know [English and new cultural ideas], because it depends on influence to be better my brain, to think and to search and to know about them."

4.1.5 RQ 1.5 What are typical English reading proficiency profiles among adult immigrant English language learners in Canada?

English reading proficiency was another source of substantial individual variation among participants in this study. Despite aiming to recruit intermediate proficiency participants at Canadian Language Benchmark levels of 5 to 8, some lower and some higher proficiency learners also chose to participate in the study, resulting in the wide range of language proficiency among participants. Among the 102 participants, the median raw total score was 16 out of 38 (42%) with a range of 4 to 35.

As discussed in the methodology, several English reading subskills were identified on the test using a cognitively diagnostic model (CDM), mastery estimates for six of which were reported to the study participants. The skill mastery estimates reported to participants were for the reading subskills of: Using vocabulary, Using explicitly stated information, Using implicitly stated information, Inferencing, Distinguishing ideas, and Using cultural knowledge (pragmatics). The relatively harder skills for study participants were 'Using implicitly stated information' (mean $PPM = .18$, $SD = .31$), 'Making connections' (mean $PPM = .22$, $SD = .33$) and 'Using vocabulary' (mean $PPM = .27$, $SD = .40$), while the relatively easier skills were 'Distinguishing

ideas' (mean $PPM = .43$, $SD = .38$), 'Using culture' (mean $PPM = .36$, $SD = .43$), and 'Using directly stated information' (mean $PPM = .35$, $SD = .39$).

Based on the probabilities of mastery for each of the six identified English reading skills, participants' typical English reading proficiency profiles were identified using latent profile analysis. Models were developed using the probabilities of mastery for each of the six skills identified by the cognitive diagnostic model: using vocabulary, using explicitly stated information, using implicitly stated information, inferencing, distinguishing ideas, and using cultural knowledge.

Several models were developed using the six skills, with different numbers of classes used in each model. The best fitting model had five classes (adjusted BIC = -462, loglikelihood value = 260, $N = 102$) and an entropy of 1.00. However, models with six classes and over did not converge, likely due to insufficient sample size. As a result, it is possible that more classes were observable, but were not identifiable in this relatively small dataset. Table 31 compares the models. Figure 7 shows, for each of the five classes in the best-fitting model, the mean probability of mastery for each reading skill.

Table 31

Probabilities of skill mastery: latent class profile models comparison table

Number of classes	Adjusted BIC	Loglikelihood	Entropy	N
2	-77	52	.99	102
3	-277	158	1.00	102
4	-394	221	1.00	102
5	-462	260	1.00	102
6 **				102

** Possible non-identification of model. The chosen model is in bold type and dark grey highlight.

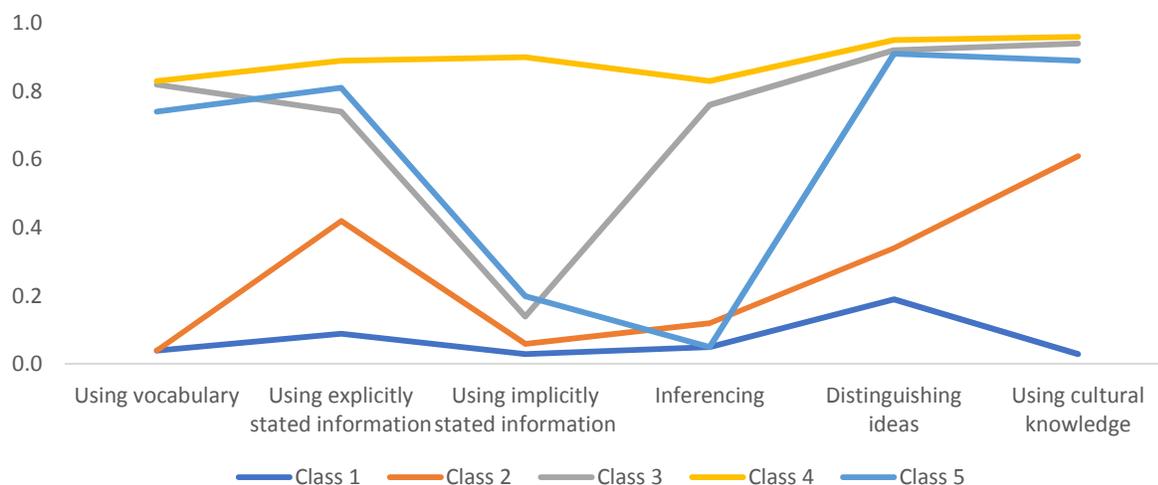


Figure 7. The mean probability of mastery on each English reading skill, for each class

Two classes had relatively low probability of mastery profiles. The largest class ($n = 61$) was Class 1, and this class received on average very low probabilities of mastery on all skills. The highest mean skill mastery for participants in this class was for distinguishing ideas, with a mean probability of mastery of .19. Class 2 ($n = 10$) was another relatively low probability of mastery class. Members of this class had on average a strong probability of mastery for using cultural knowledge (mean $PPM = .61$, $SD = .18$), and a moderately low probability of having mastered using explicitly stated information (mean $PPM = .42$, $SD = .25$). Other skills were associated with very low probabilities of mastery.

Two additional classes had high probabilities of having mastered using vocabulary, using explicitly stated information, distinguishing ideas, and using cultural knowledge, but mixed results for the skills requiring use of indirectly given information. Class 5 ($n = 9$) had low mean probabilities of mastery for both using implicitly stated information (mean $PPM = .20$, $SD = .15$) and inferencing (mean $PPM = .05$, $SD = .07$), while Class 3 ($n = 7$) had a high mean probability of having mastered inferencing (mean $PPM = .76$, $SD = .18$) but a low mean probability of having mastered using implicitly stated information (mean $PPM = .14$, $SD = .14$).

Finally, one class (Class 4, $n = 15$) had high mean probabilities of mastery for all six skills, with the joint lowest mean probability of mastery for using vocabulary and inferencing at .83 ($SDs = .31$ and $.20$ respectively).

Overall, the raw total score was associated a gradual increase in the probabilities of mastery for the six English reading skills reported on the report. The latent profile classes derived from the probabilities of mastery can be ordered by proficiency in this order (least proficient first): Class 1, Class 2, Class 5, Class 3, Class 4. Correspondingly, the mean raw total score for participants in Class 1 was 12.56 ($SD = 4.16$), for Class 2 was 19.10 ($SD = 2.60$), for Class 5 was 24.33 ($SD = 2.74$), for Class 3 was 26.71 ($SD = 2.81$), and for Class 4 was 30.53 ($SD = 2.80$).

4.1.6 RQ 1.6: How do adult immigrant English language learners in Canada perceive their English reading proficiency?

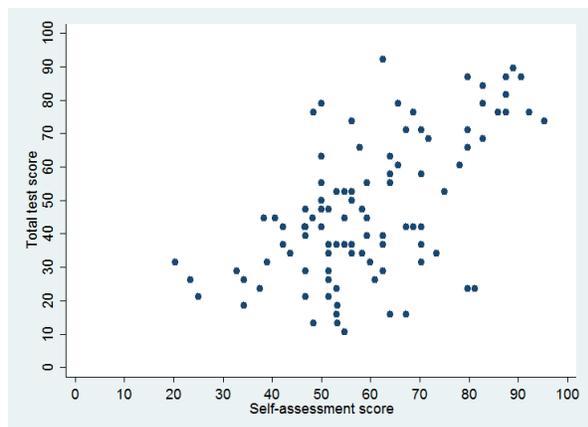
Alongside the variation in English language proficiency as reported by the test results, participants' own self-assessments showed variation. On a scale of 1 to 5, participants self-assessed to a mean of 3.37 ($SD = 0.63$), and individual self-assessment scores ranged from 1.81 to 4.81. There was also a reasonably strong relationship between total test score and self-assessment score, with a Pearson's r correlation of .59 ($p < .000$), indicating that participants less proficient in English reading (as per the test results) self-assessed relatively lower than higher proficiency participants.

However, a participant who scored a low test score was unlikely to self-assess at an equally low level; in general, these participants overestimated their proficiency compared to the test results. In contrast, the higher-proficiency participants were much more likely to self-assess at a level comparable to their performance. This relationship is visualized in Figure 8 and Figure 9.



*Test and self-assessment scores have been placed on a 0 to 100 scale for comparability

Figure 8. Plot for test score vs. perceived score in this study



Test and self-assessment scores have been placed on a 0 to 100 scale for comparability

Figure 9. Relationship between total raw score and overall self-assessment score

Next, participants' self-assessment scores were placed with participants' raw total test results into a latent profile analysis to identify whether groups of participants emerged who substantially over- or underestimated their English reading proficiency. The best fitting model had four classes (adjusted BIC = -184, loglikelihood value = 102, $N = 102$) and an entropy of 0.80, indicating adequate classification quality. Models are compared in Table 32.

Table 32

Self-assessment discrepancy: latent class profile models comparison table

Number of classes	Adjusted BIC	Loglikelihood	Entropy	N
2	-179	95	.84	102
3	-182	99	.77	102
4	-184	102	.80	102
5	-184	104	.78	102

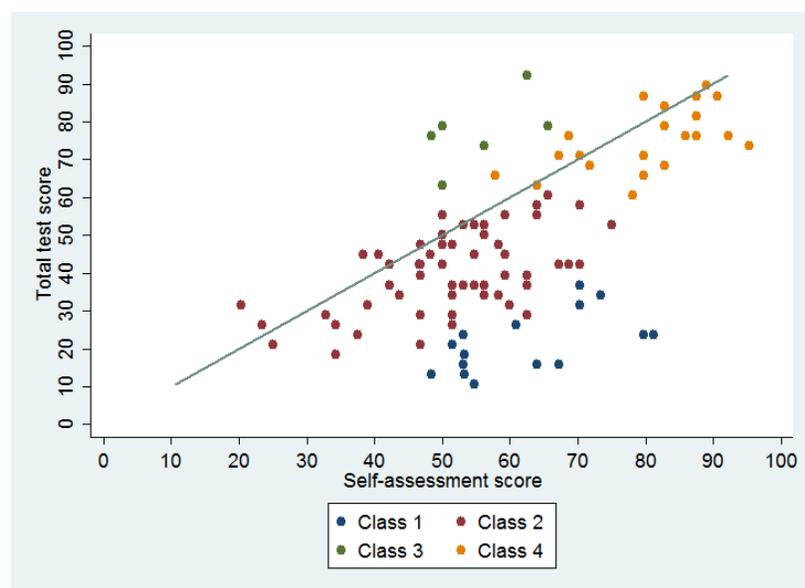
** Possible non-identification of model. The chosen model is in bold type and dark grey highlight.

As expected, the four classes were associated with extent of over- and under-estimation, and this association is depicted in Figure 10. Class 1 ($n = 15$) consisted of participants who drastically overestimated their English reading proficiency. These participants were characterised by relatively low raw test scores and a very large discrepancy. Members of Class 1 had a mean raw test score of 8.20 out of 38 ($SD = 3.05$) but a mean self-assessment score of 3.49 on a scale of 1 of 5 ($SD = 0.43$). This corresponds to a mean 21.58% of items correct ($SD = 8.03$ percentage

points), but a mean self-assessment of 62.32% ($SD = 10.81$ percentage points), resulting in a mean discrepancy of +40.74 percentage points.

Two classes consisted of participants who either accurately estimated or moderately overestimated their English reading proficiency. Class 2 ($n = 59$) contained relatively lower English reading proficiency participants. In this class, the mean raw test score was 15.20 out of 38 ($SD = 3.97$, mean percent correct = 40.00%, $SD = 8.88$ percentage points) and the mean self-assessment score was 3.05 on a scale of 1 of 5 ($SD = 0.46$, mean percent correct = 51.31%, $SD = 11.46$ percentage points). Class 4 ($n = 22$) contained higher proficiency participants, and the mean raw test score was 28.41 out of 38 ($SD = 3.23$, mean percent correct = 74.76%, $SD = 8.50$ percentage points) while the mean self-assessment score was 4.20 ($SD = 0.40$, mean percent correct = 80.06%, $SD = 10.02$ percentage points).

Finally, Class 3 ($n = 6$) consisted of several high English reading proficiency participants who all underestimated their proficiency. Participants in this class had a mean score of 29.33 ($SD = 3.56$, mean percent correct = 77.19%, $SD = 9.37$ percentage points) yet a mean self-assessment score of 3.22 ($SD = 0.29$, mean percent correct = 55.47%, $SD = 7.25$ percentage points), resulting in a mean discrepancy of -21.72 percentage points.



*Test and self-assessment scores have been placed on a 0 to 100 scale for comparability

Figure 10. Relationship between total raw score and overall self-assessment score, with latent profile classes

4.1.7 RQ 1.7: What relationships are observable between individual characteristics?

In order to open a window into the network of relationships that existed among the individual characteristics observed in this study, a correlational analysis was conducted for those characteristics observed in the form of quantitative variables. In terms of basic demographic characteristics, it is important to note that there were no gender or education differences for other background variables with the exception that participants with higher levels of education often had higher levels of English reading proficiency ($r = .42, N = 100, p < .001$) and knew it ($r = .27, N = 100, p = .006$). Additionally, women were somewhat less likely to use English in emails regularly ($r = -.23, N = 101, p = .020$), or feel confident about being able to read English language novels in future ($r = -.20, N = 97, p = .045$). Participants' age was relevant to English reading proficiency; younger participants were likely to be more proficient in English reading ($r = .40, N = 94, p < .001$) and know it ($r = .29, N = 94, p = .005$). Younger participants were also more likely to expect to make friends, read novels, and watch TV in English. These age-related findings are because in this study, about 20% of participants were international students studying in high school, college or university, and already had relatively high levels of English proficiency. As a result, they generally scored higher on the test and were confident about their English skills.

Additionally, younger participants were somewhat more likely to use English in online social media ($r = .27, N = 93, p = .010$) and for internet-based conversation using software such as Skype ($r = .26, N = 94, p = .013$). This finding likely reflects a generation gap in usage of digital platforms and products. Similarly, younger participants were somewhat more likely to use English to talk with friends ($r = .23, N = 94, p = .028$). This finding is explained by that hypothesis that as younger immigrants reported having more time to socialise and make friends outside their immediate community, they are more likely to have friends in their peer group who come from other language backgrounds. This is particularly true for those participants in full time study contexts such as high school, college and university. Note that background characteristics were not related to psychological characteristics, which is in accordance with theoretical expectations.

In terms of proficiency in languages other than English, this study did not delve deeply, as noted in the Methods chapter. The language profiles of this study's participants were highly diverse,

with participants reporting knowledge of 31 languages, 24 first languages, and 18% of participants reporting knowledge of at least two languages in addition to English. Moreover, reported typical usage and functional abilities for known languages, including use of English at home in place of other known languages, also varied widely, with no consistent observable profiles. As a result, there was little opportunity to conduct analyses according to language profile. Correlational analyses were conducted to identify whether reporting a Chinese ($n = 39$), Romance ($n = 14$), or Indo-Iranian ($n = 21$) first language was associated with any other background variables, but no significant associations were observed.

The variables for participant psychological characteristics yielded interesting relationships. While stronger fixed beliefs about intelligence were associated only with stronger performance avoid goal orientations ($r = .32, N = 97, p = .001$), goal orientations had consistent relationships with English usage, English reading proficiency, perceived proficiency, and confidence in future English abilities. First, strength of mastery goal orientation was not associated either higher or lower English reading proficiency, but participants with stronger mastery orientations were likely to believe they were more proficient ($r = .30, N = 101, p = .003$), and perhaps unsurprisingly therefore, also had greater confidence in eventually being able to use English to make friends ($r = .21, N = 98, p = .039$), read novels ($r = .23, N = 97, p = .021$), watch TV ($r = .24, N = 97, p = .018$) and get a job ($r = .32, N = 97, p = .002$). Participants with a stronger mastery goal orientation were also somewhat more likely to use English orally, for shopping ($r = .21, N = 101, p = .038$), watching TV ($r = .21, N = 101, p = .033$), and talking to friends ($r = .23, N = 101, p = .020$), as well as on social media ($r = .27, N = 100, p = .007$).

In contrast, participants with stronger performance prove orientations were more likely to regularly use English to talk to family ($r = .22, N = 101, p = .028$), but there were no further significant relationships observed with either English usage, English reading proficiency or confidence for future English proficiency. Likewise, English usage was generally unrelated to extent of performance avoid orientation, except that participants with stronger performance avoid orientations were somewhat less likely to regularly watch TV in English ($r = -.23, N = 101, p = .019$). However, they were also somewhat more likely to have lower English reading proficiency ($r = -.21, N = 101, p = .039$), know it ($r = -.26, N = 101, p = .008$), and have less faith in their future English skills ($r = -.24$ to $-.33, N = 97$ to $98, p < .05$). Their lesser

engagement with English TV may simply be due to their lower English language proficiency. Note however (as mentioned previously), that a stronger performance avoid orientation was also associated with stronger fixed beliefs about intelligence, indicating that underlying beliefs about learning potential may be affecting these participants' progress.

The relationships between self-assessment and English usage were also very interesting. A higher self-assessment was clearly related to higher test score ($r = .59, N = 101, p < .001$), indicating at least moderate awareness of English reading proficiency among participants. As noted previously, a higher self-assessment was also associated with higher mastery orientation and lower performance avoid orientation. In addition to these findings, more frequent English language usage was moderately associated with a higher self-assessment. In particular, participants with higher self-assessment scores were more likely to read English books regularly ($r = .40, N = 101, p < .001$) and somewhat more likely to use English on the internet regularly (social media usage $r = .21, N = 100, p = .033$; online chat $r = .26, N = 101, p = .008$) and to talk to friends in English ($r = .23, N = 101, p = .019$). However, the frequency of these activities was not related to reading test score.

In contrast, two activities that were positively related with both higher test scores and higher self-assessment scores were using English to shop (test score $r = .35, N = 101, p < .001$; self-assessment score $r = .34, N = 101, p < .001$) and write emails (test score $r = .38, N = 101, p < .001$; self-assessment score $r = .29, N = 101, p = .003$). Note that these two activities require communicating with strangers and writing – perhaps also to strangers, respectively. This is in contrast to the activities that are associated with higher self-assessment but not higher test score. An explanation is that these activities provide more realistic feedback and language/reading skills than the informal, sympathetic, or isolated contexts above.

One additional finding is that participants with higher English reading proficiency were somewhat less likely to use English with their family ($r = -.25, N = 101, p = .014$), likely because younger participants, who were less likely to have children (a major source of within-family English usage) had generally higher English reading proficiency in this study.

Finally, participants' confidence in their ability to function in English, for example by making friends in English, reading English language novels, watching TV in English and/or getting a job that uses English, was consistently related to mastery and performance avoid orientations, beliefs

about intelligence, and English reading proficiency and perceived proficiency. As previously noted, English reading proficiency and perceived proficiency were both positively associated with confidence in being able to do all four activities; a logical outcome given that participants with sufficiently high proficiency would already be able to do some or all of these activities (or believe they can).

Likewise, a higher mastery orientation was associated with somewhat greater confidence in being able to do these activities in future. In contrast, a performance avoid orientation was moderately associated with weaker confidence. In addition, stronger fixed beliefs about intelligence were associated with weaker confidence in future ability to read novels ($r = -.30, N = 93, p = .003$) or watch TV in English ($r = -.30, N = 93, p = .003$).

4.2 Theme 2 – Attention

4.2.1 RQ 2.1: To which aspects of feedback reports do language learners report paying attention?

The first set of results describes the attention that participants reported giving to each area of the report. Immediately after receiving their report and completing the planning section, all participants were asked on a survey how much they looked at each section of the report: the introduction, the skill descriptions, the figures comparing test and self-assessment scores, the learning suggestions, and the planning opportunity. They were invited to respond on a 1 to 5 scale of ‘I did not look at this’, ‘a little time’, ‘some time’, ‘a lot of time’, and ‘all my time’. (Note that participants who selected ‘all my time’ for a section also reported spending time on other sections, a logical inconsistency to be avoided in subsequent research through usage of scales of degree). Figure 11 shows the percentage of participants selecting each response for each section of the report.

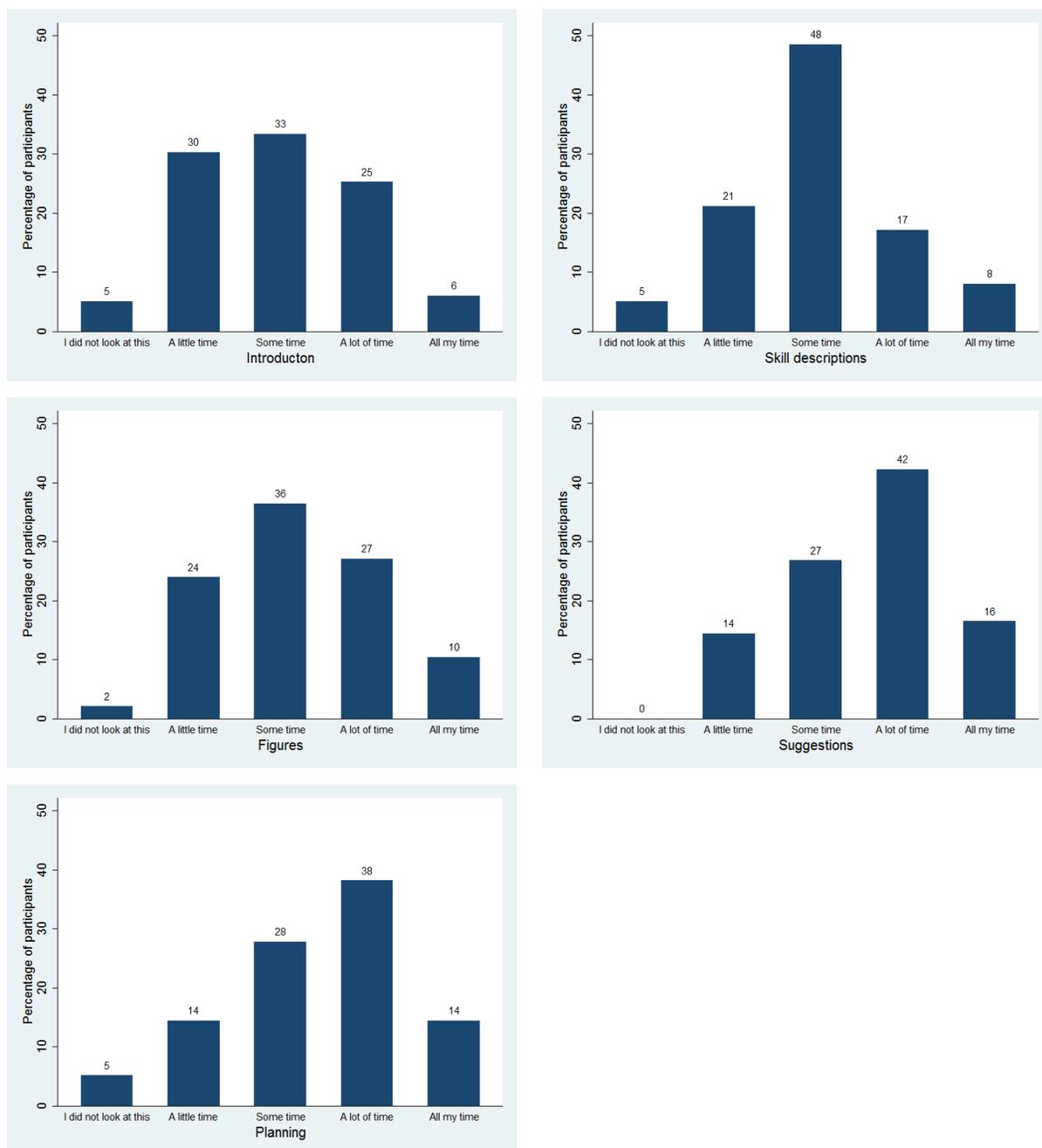


Figure 11. Percentage of participants reporting amount of time spent looking at each section of the report.

The suggestions for learning were reported to be looked at the most, with 58% of participants reporting that they spent ‘a lot of time’ or ‘all his/her time’ on this section, and only 14% reporting that they spent ‘a little time’ looking at this section. The suggestions were also the only section in which everyone reported spending at least some time. The planning section also received a lot of attention with 52% of participants reporting that they spent ‘a lot of time’ or ‘all

his/her time' on this section, and only 19% reporting that they spent 'a little' or no time looking at planning.

The section receiving third-most attention was the figures that compared test and self-assessment scores. Thirty-seven percent of participants reported spending 'a lot of time' or 'all his/her time' on this section, although overall the responses were fairly normally distributed, with a small negative skew; 26% of participants reported spending 'a little' or no time on this section. Kolmogorov-Smirnov tests indicated the distribution was significantly different from that of the suggestions section ($D = .21, p = .021$).

In contrast, the distribution of reported time spent on the skill descriptions was almost fully normal, indicating the participants reported no strong tendency to much or little attention on this section. In fact, very few participants reported spending a lot or only a little time on the skill descriptions, and Kolmogorov-Smirnov tests indicated the distribution was significantly different from those of the suggestions and planning sections ($D = .34, p < .001$ and $D = .27, p = .001$ respectively). Just 25% of participants reported spending 'a lot of time' or 'all his/her time' on the skill descriptions and 26% reported spending 'a little' or no time. Nearly half (48%) of participants reported simply spending 'some time' looking at the skill descriptions.

Reported time spent on the introduction was also fairly normally distributed, but participants were more divided regarding the time they spent on it; 31% of participants reported spending 'a lot of time' or 'all his/her time' on the introduction compared to 35% who reported spending 'a little' or no time. Like the skill descriptions, Kolmogorov-Smirnov tests indicated the distribution was significantly different from those of the suggestions and planning sections ($D = .27, p = .001$ and $D = .21, p = .012$ respectively).

In addition to the survey, 12 participants also completed interviews in which they were shown eye-tracking traces of their report reading, and asked to recall their thoughts. The depth and extent of cognition reported by participants consistently aligned strongly with the amount of time they reported looking at each section of the report. Four examples of reported attention patterns highlight the consistent alignment between reported amount of time on a section, and reported cognition for that section.

For example, participant EL reported spending only ‘a little time’ on each of the report sections. In her interview, EL consistently reported focusing on reading through the report in good time rather than understanding the report content. When asked what she was thinking as she was reading, EL repeatedly said ‘I didn’t think anything! ...No, I didn’t think when I read here...Just read and then, yeah, then finish...This sentence here I just maybe have, I think, oh that’s very good. I can’t think too much because I just want to go through the sentence.’ Note that it later became clear that EL (mistakenly) thought she was under timed conditions similar to writing a test, and additionally EL was in fact the participant who subsequently reported most usage of the report in interviews a month later. Therefore, while her reported time in the survey and interview matched up well, additional factors explained her subsequent report usage.

In contrast to EL, participant UB reported spending ‘all her time’ looking at the skill descriptions, graphs and suggestions, and ‘a lot of time’ looking at the planning section. Like EL, UB reported spending ‘a little’ time looking at the introduction. UB reported detailed attention to the skills, figures, suggestions and planning, demonstrating her interest in all the details: “I want to understand fully about the next part because I know that it will tell me how my work was done and to evaluate my total skills so I wanna understand it 100% to get a better understanding for the next part... I was just thinking the material to tell me the problem about my vocabulary. I wanna figure out how does it work. Because this is the first time I see the graph and the words to put like this, so I wanna understand, I wanna find some advice for me and some, like, problem of my skills...”

In addition, participants BO and DP reported distinctly uneven amounts of time spent across report sections, with BO initially reporting a lot of time and then dropping off, while DP reported increasing time given to the sections as he read his report. Specifically, BO reported spending ‘a lot of time’ on skill descriptions and ‘all his time’ on the figures that compared his test results with his self-assessment, but only ‘some time’ on suggestions and ‘a little time’ on planning. His reported cognition for skill descriptions and figures is detailed and questioning: “Just big difference between test says and you say, so. What I say? (laugh) I was just wondering why there is such big gap between the two... I think I was thinking, what is direct stated information is exactly. So I check the former one, using vocabulary, but there's no relations between both so I just um, checked the title again using directly stated information. I want know what it says... Just

reading, thinking, reading, thinking.’ In contrast, BO reported substantially less engagement in suggestions and planning: “I, usually the first, first section paragraph, section. There's no, nothing spe, I mean, it's mention general things so I like to read second paragraph first. And then I just, I just um, I just skimmed the first lines... I didn't read those uh, line clearly. I was, I, I rather think then read, so. And, it's so general, uh yeah, I know that, everyone knows that, so. Nothing special, yeah, we all know that (laugh).”

Conversely, DP reported spending ‘a little time’ looking at the introduction and ‘some time’ looking at the skill descriptions, but ‘a lot of time’ looking at the figures, suggestions and planning. His reported cognition bears this out, with DP clearly valuing the introduction and skill descriptions less than other report sections. For example, DP’s reported cognition for the introduction was represented in statements such as “No I didn’t, I just look through... Because it’s the covers I don’t, like, think it’s important so I just quickly read it, yeah.” Likewise, DP rarely commented on skill descriptions, and when he did his comments were minimal, for example “I just read and read... I usually don’t, like, pay too much attention to the word, yeah. I’m a picture person.” In contrast, DP reported substantial cognition around the figures, for example: “The test said, because I want to see what the test says, yeah. Because the test is more, like, more, it’s real. It’s like, not, without, like, my personal opinion on this thing... just thinking about why the result is like that. Yeah.” DP also engaged with the suggestions and planning, for example noting the suggestions’ value and critically evaluating them “I think maybe it’s the most important part for me, and I can get some good information and helpful information from this page. So I take this page seriously... This something in the textbook, like, I buy a lot of reading textbook, it always say that thing, the writer’s feeling you are writing, so, yeah. It’s very typical. It’s useful but I know. I already knew that.”

In summary, these results indicated that the amount of time that participants reported spending looking at each area of the report was strongly related to the amount of cognition and effort that participants put into each area of the report.

4.2.2 RQ 2.2: To which aspects of feedback reports are language learners observed to pay attention?

The second set of results describe the attention that participants were observed to give to each area of the report. Eye-tracking traces for participants who completed eye-tracking interviews

were available to offer empirical evidence about areas of the report that learners were observed to pay attention to. Of the fifteen participants who completed the interviews, eleven had eye-tracking trace data of sufficient quality to use, and enough language proficiency to be able to comprehend the report. Fixation records for these participants indicated which areas of the report they looked at most, and included duration of fixations in each area, and number of fixations in each area.

In contrast with the reported time spent looking at report sections, the fixation data indicated that on average, participants spent approximately equal amounts of time looking at the skill descriptions, suggestions and planning – about 25% of their time on each section, with slightly more on skill descriptions. While overall participants reported spending a lot of time looking at suggestions and planning – ranking first and second in amount of reported time, skill descriptions placed joint-fourth (last) rank for amount of reported time. However, the fixation data indicate participants in fact spent a lot of time looking at the skill descriptions.

Similarly, the fixation data indicated that on average participants spent only 5% of their time looking at figures, a clear fifth rank for amount of observed attention. However, this observation is in direct contrast to the amount of time reported for this section; 75% of participants reported spending ‘some’ to ‘all their’ time on this section and it ranked a close third in amount of reported attention. The only report section in which observed time resembled reported time, was the introduction. The amount of time spent on the introduction was on average 10% of total time, which placed it in fourth rank for amount of attention, comparable to the reported amount of time, in which it also placed joint-fourth rank.

Overall, these findings indicate that amount of time spent on a section was most strongly related with the amount of text in the section rather than providing information on cognitive attention. The skill descriptions, followed by the suggestions and planning, were the most text-heavy section. The figures had very little text.

There is also some evidence from the fixation data that participants spent relatively more time fixating on descriptions and figures of skills for which they received lower mastery estimates. This finding was observable both among participants with generally higher mastery estimates but some low estimates, and those with consistently low but non-zero estimates. For example, participant ME received skill mastery estimates of less than .10 for all skills except Using

directly stated information, which had a .14 probability of mastery. ME devoted 4% to 8% of his time to each of those very low probability of mastery skills, but only 3% of his time to Using directly stated information. Similarly, participant ML had probabilities of mastery of less than .05 for all skills except Using directly stated information, which had a .54 probability of mastery. Accordingly, ML devoted 4% to 6% of her time to each of the skills with the exception of Using directly stated information, which she only fixated on for 2% of her time. At the other end of achievement, participant BO had probabilities of mastery of .94 and over for all skills except Using implicitly stated information and Inferencing. Reflecting this profile, he spent 3% to 5% of his time fixating on each of the skills with high probabilities of mastery, and 8% of his time on each of his two relatively weak skills. However, other factors such as order of presentation, amount of self-assessment discrepancy, and other unexplained influences also appeared to exert influence on time spent on each skill.

The eye-tracking traces also provided information about the order in which participants viewed the report sections, and how often they visited them. Note that due to limitations in the eye-tracking technology used in the study, scrolling was not possible during eye-tracking recording, and the text needed to be sufficiently large that line-by-line tracking could be observed. As a result, the report was split into six pages for viewing via the eye tracker: introduction; VOC and EXP skill descriptions and figures; IMP, INF, DIS and PRG skills descriptions and figures; suggestions for learning; planning introduction and first planning text box; other two planning text boxes. It was also not possible to enter text via the eye-tracking device – participants did it later. Based on these limitations, several observations were possible. Between pages it is mainly possible to observe variation in amount of attention, as discussed above. Within pages it is also possible to observe where participants looked, in what order, and how often. It is this set of within-page observations that is now described.

For within-page analysis the most interesting pages were the two containing figures and skill descriptions. Whereas the other pages were fully text and participants displayed regular reading patterns expected of language learners, eye-tracking traces showed that participants alternated between skill titles, bullet points, and figures in various ways.

Firstly, for five of the seven participants who had very low skill profiles – profiles with near-zero probability of mastering each skill, fixations on the figures were minimal (see Figure 12),

although they read through the skill descriptions and titles consistently. Note that for these participants, the figures provided little information other than to tell the test taker that they scored very low on the test, perhaps explaining this absence of fixations.

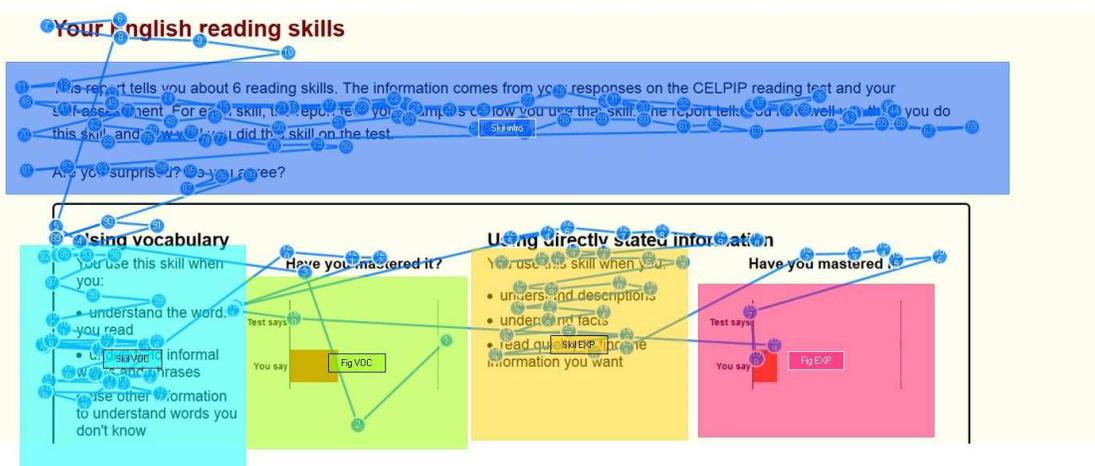


Figure 12. Eye-tracking traces for a participant with very low skill mastery

Among those with mixed and higher mastery profiles there were more fixations on the figures, although with substantial variation. However, a consistent trend was focusing fixations on the test results bar. For example, participant BK received a very mixed skill profile that included both over estimates, under estimates, and accurate estimates. Regardless of the discrepancy, BK fixated on both the very low and very high test results bars, and few fixations were recorded for the self-assessment bars (Figure 13). Similarly, BO had generally high estimates of skill mastery and substantially underestimated both his first two skills, but fixated almost entirely on the test results bars (Figure 14). The results bars appear to be operating as a replacement for more traditional test scores, and become a primary point of reference for participants with non-zero results.

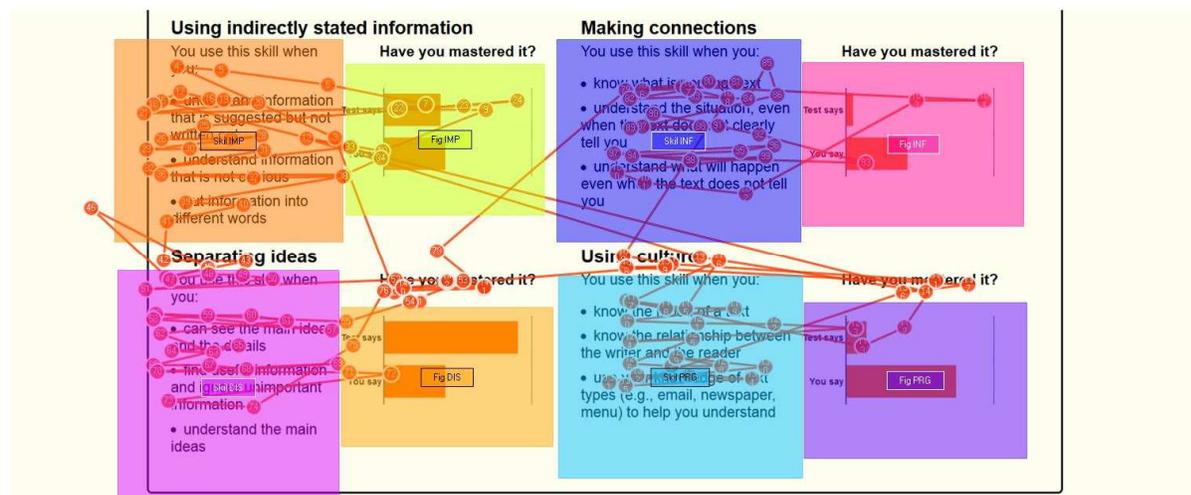


Figure 13. Eye-tracking traces for a participant with mixed skill mastery: BK

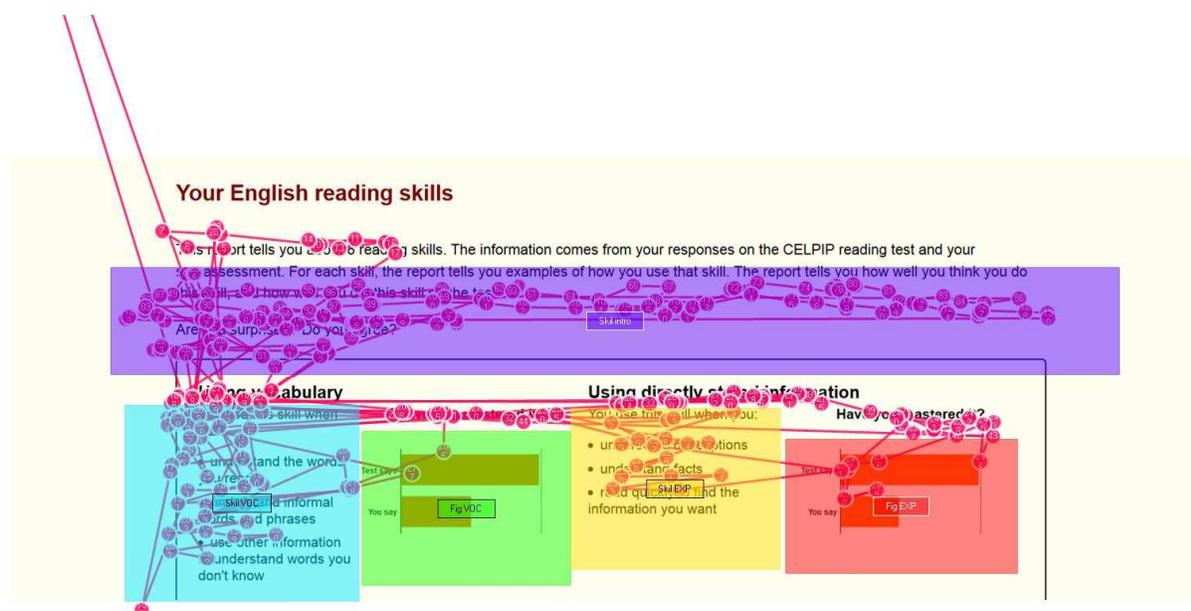


Figure 14. Eye-tracking traces for a participant with high skill mastery: BO

A third feature to note is the variation in attention to skill titles. Skill titles, operating as names, are important for being able to recall skills that were tested. However, fixation traces indicate that for thirteen of the fourteen participants with usable eye-trace data, the titles of the first two skills (Using vocabulary and Using directly stated information) were consistently fixated on, while fixations on the third and fourth skills (Using indirectly stated information and Making connections) – which were placed at the top of the next page – were often non-existent. Fixations were generally found on the fifth and sixth skills (Separating ideas and Using culture) as participants continued reading from the bullet points above. Figure 15 presents an example of the differences in fixations for skill titles. This finding indicates that once participants had oriented themselves to the structure of the report, they did not use the skill titles to orient themselves or to make meaning of the bullet points below. In contrast, there was consistent interest in looking at the bullet points, and it is hypothesized the purpose of this is to make meaning of the results the participants received on each skill.



Figure 15. Eye-tracking traces showing uneven fixation distribution on skill titles

4.2.3 RQ 2.3: Which aspects of feedback reports can be recalled by language learners one month later?

In addition to eye-tracking interviews at the time of receiving the report and the survey that all 102 participants completed, 15 participants completed delayed recall interviews, during which one month after receiving their report, they recalled what they remembered of the report and discussed how they had been using it. Without prior reminding of their report content, participants were asked the following six questions:

1. How have you used your reading report?
2. What do you remember about your reading report?
3. What did the report tell you about your reading skills?
4. What did the report look like?
5. What did your report suggest you should do to continue learning?
6. What learning goals did you set?

This data provided empirical information about which aspects of the report to which they sufficiently paid attention that they recalled without prompting one month later. A hierarchy emerged regarding which areas of the report participants recalled. The figures were clearly the most remembered part of the report. Of the fifteen participants, 12 recalled the figures, including 11 who mentioned their results and the discrepancies between the test results and their self-assessment – which could only be derived from the figures. Comments referring to the discrepancies noted the differences, for example “I remember there is a, I think, graphic that said I’m weak in this or I get for example less maybe in reading. And it was some marks good and some marks weak”, or interpreted the differences as low achievement, for example “My reading skill is bad!” or “Some report tell me the reading is I think more poor because I understand a bit a little something.” The three who failed to mention the figures or results recalled very little at all from the report.

The next most-recalled areas of the report were the planning section and the suggestions for learning. Four participants recalled only the figures and the plans, two participants recalled only the figures and the suggestions, and one participant recalled the figures, suggestions and plans from his report. The suggestions recalled were ones that participants found personally meaningful, for example “The report told me how can I improve my reading skills. Read the

newspaper or something else, and listening to radio and watching TV. And I think work. Working in English situation...I remember first and second advice exactly” and “Read some label, about label something, label, so sometime I will, when I went to the store I will read the label.” More specific suggestions regarding habits while reading, for example predicting what will come next, were not recalled by any participants.

The specificity with which plans were recalled varied, and were related to the specificity of the original plans. For example, the participant who recalled “Yeah so I want to know about the vocabulary...” originally wrote “I don't understand many difficult vocabulary. So I must study it. Every day I will read a newspaper and check difficult words.” In contrast, the participant who recalled “I say should continue taking our classes here, LINC classes, and maybe listening news, talking in the communities, maybe finding networking to have practice for conversation. I believe that I wrote that doing homework with my kids helps me too, so maybe I wrote something like that” originally wrote “I will help my children to study at schools-with their homework; I will continue taking my LINC classes; I will widen my communication with native speakers; I will continue any kind of reading, listening, watching and using.”

Finally, three participants recalled something about the skill descriptions, mainly the names of the skills, although no one recalled all six skill names. One participant recalled some of the bullet points in the descriptions. All three of these participants also recalled the figures, suggestions and plans, and reported using the report in some detail. It appears that recall of the skill descriptions is at the bottom of an inverted hierarchy of recall, with engagement in other areas of the report required before skill descriptions are processed sufficiently to be recalled one month later.

The inverted hierarchy of recall for the report sections was therefore figures, plans and suggestions, and skill descriptions, shown in Figure 16. None of the participants mentioned the introduction. Recall of the figures or the information contained in the figures appeared to be a prerequisite for being able to recall anything from the report. Plans and suggestions were recalled in different ways, and finally descriptions could be recalled if there was enough investment in the content of the report.

This inverted hierarchy indicates that the extent of personalization in a section affected the long term recall of that section. The figures contained highly personal information, including participants' test results, their self-assessment results, and a comparison of the discrepancies

between them. Clearly this level of personalization and personal significance had an impact. The plans, which were made by the participants themselves, were also personal and additionally were self-determined, likely facilitating long term recall. The suggestions that were deemed by participants to be relevant to their lives were also likely to be recalled. In contrast, the skill descriptions were only recalled to the extent that they provided meaning to participants' understanding of the report, and were only recalled by participants who had clearly devoted substantial attention to the report. Finally, the introduction, which was generic with the exception of the participants' name, was not mentioned at all.

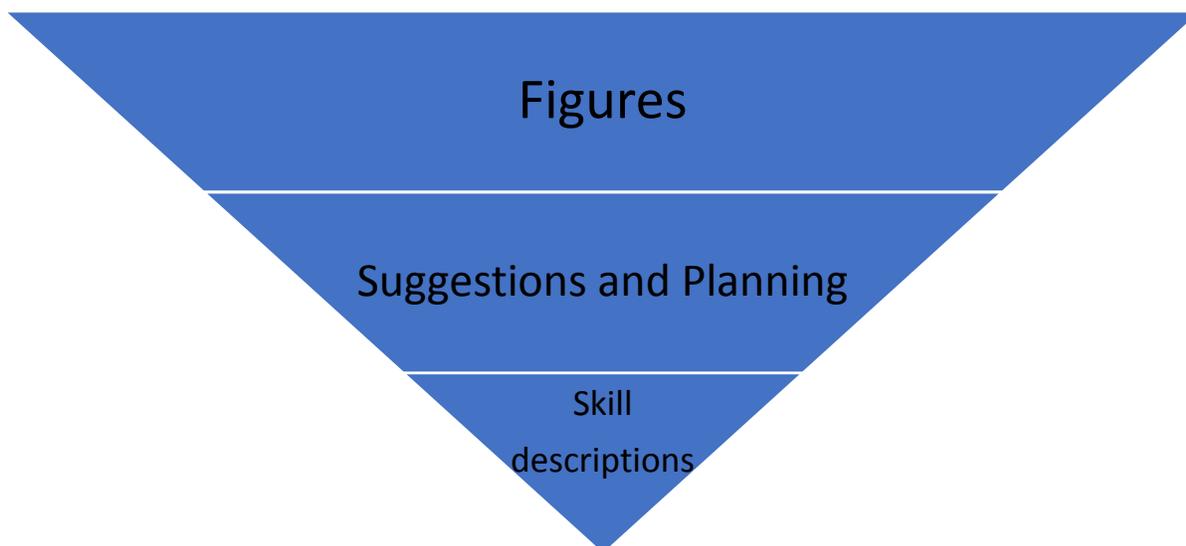


Figure 16. Hierarchy of recall for report sections

4.3 Theme 3 –Interaction with report content

4.3.1 RQ 3.1: What affective and cognitive strategies do language learners report when receiving a feedback report?

Eleven participants' eye-tracking interview data was used to identify how participants interacted with the report. Through coding themes as they emerged from the interviews, and with reference to the self-regulated learning literature, five themes were identified as arising from the interviews, each of which had two or three categories of behaviour observed among participants. The themes were broken down into two main self-regulated learning categories (Pintrich, 2004): Affective – dealing with negative results, emotionality of response to the report, and Cognitive – extent of relating report content to one's own life, dealing with comprehension problems, and type of critical evaluation of report content. The findings, including variations in behaviour observed

among participants, are summarized in Table 33. The observed variations in behaviour for each these topics are discussed further below.

Table 33

Eye-tracking interview participant interaction profiles

Participant	Dealing with negative results	Emotionality	Relating report to own life	Dealing with comprehension issues	Type of critical evaluation **
1	minimized		some	want help	positive only
2		emotional	no	none	positive and negative
3		emotional	lots	self-resolving	positive and negative
4	minimized		lots	want help	positive only
5			some	self-resolving	positive only
6	minimized		no	want help	positive and negative
7	minimized	emotional	some	want help	none
8	minimized	emotional	no	self-resolving	positive and negative
9			some	want help	none
10	minimized		lots	none	none
11		emotional	no	want help	none

** 'Positive' and 'negative' refer to conclusions drawn as a result of critical evaluation

4.3.1.1 *Dealing with negative results*

All participants received either flat non-mastery profiles or mixed profiles containing high and low probabilities of mastery. In addition, all participants overestimated their skills on at least one skill. As a result, all participants had to deal with negative results – either low probabilities of skill mastery and/or discrepancies between their self-assessment and test results. Some participants dealt with these unexpected results by minimizing their personal responsibility for those results.

In total, six participants minimized their personal responsibility for the unwanted results, and they did this by explaining the results as due to external or extraneous factors outside of themselves. Although the design of the test requires timed responses to create meaningful results, and reading speed a known aspect of language proficiency, two participants believed that they

were given insufficient time to demonstrate their reading skills, for example: "...but the test, the time factor is strong. Maybe if we did it all, eh, half more time, is different result. Yeah? But when the time is quickly, is difficult catch to many ideas." Two others referred to aspects of their personal background that could not be changed, one questioning whether there were underlying cultural factors at work and another referring to age: "I was thinking, is it a problem with culture, or with understanding special words? ...Why is there such a big discrepancy? Maybe it's cultural differences...this is something written in English, but my mind uses Chinese!" [Translated from Chinese.] "But not much, practise this...when I was older this is not much practice. Usually younger [people learn more easily], right?" One other participant said she was tired on the day of the test and that the result was not a surprise to her. Finally, one participant explained the discrepancies by stating she misunderstood the self-assessment parameters: "But I, I, I gave the answers on all, uh, skill and situation and what I feel how much I understand...in general... what is the difference between the program's rating and my rating. It should not be that much [difference]... You can, you can consider 50% of those, what I say."

Note that most of these participants did attempt to not completely explain away their lower-than-expected performance. For example, the participant above who was quoted blaming the limited time had also previously noted, in response to a low probability of mastery on one skill: "Understand the words you read...People need to use them, yeah. Erm, for me is, is easy, relative, is easy understand the words, the formal words, but the informal words is very difficult."

4.3.1.2 *Emotionality of response to the report*

While some participants reported little emotion while they received their report, five participants reported emotional responses. Many emotional responses focused on discrepancies between test results and self-assessments. An example of such a response is: "I think I very um, humble, I don't know how to say, but why the test says even, even shorter than what I say. So, I think it the bars was weird, so. (Laugh)." Another example is "I was a little bit, kind of, astonished, because my evaluation to myself is much better than the test. Than the evidence. So I was a little bit, kind of, astonished, so I want read this quickly."

Of the three participants whose reports included overestimates, only one reported emotion about an overestimate: "I was happy! Because the test give me a really confident. Because I'm not really confident about my total English levels so when I saw this I don't know if this is 100% or

something, I'm just very happy so I look at this and I compare this two." Similarly, only one participant responded to a consistently low mastery profile in emotional terms: "I thought about my result for the test...My English skill is very not good, and then, so...I'm little, little shy (laugh)...Yeah I'm disappointed, maybe at that time."

A caution for interpreting this topic is that cultural differences exist among people groups regarding the extent to which emotion is expressed. Therefore, although some participants did not appear particularly perturbed by their report, for example: "So but you see, you can see here the difference (laugh) just say, and that I'm more than (laughing) and it is the contrary. I was, curious about these, these differences", it is possible that some participants suppressed emotional responses.

4.3.1.3 Extent of relating report content to one's own life

A very common theme was that of relating report content to one's own life, in which participants engaged to varying degrees. Four participants focused entirely on the content of the report – evaluating it, comprehending it, reacting to it, and reported no thoughts that connected the report to their own lives. Four other participants made some links to their lives, when thinking aloud through the introduction and suggestions for learning. Finally, three participants consistently linked their lives with the report throughout their reflection.

Overall, the suggestions for learning prompted the most connections to participants' own lives. Examples of reflection include: "It says must tell stories, it means my kids, I am telling stories, sometimes" and:

Use cultural knowledge. Ok. Now I apply to the volunteer in the festival of documentaries and I take ah, a position...It is second time that any person say to me apply to volunteering, it's very important. I think is a very good idea.

However, the introduction also prompted all seven participants who thought about their own lives to do so. These comments were generally statements citing the necessity of acquiring English for their own lives, for example: "I have to achieve, I have to know. Like to communicate with other reading and understand newspaper in daily life. For example, I receive email from you, I have to understand by myself." Therefore, comments based on the introduction

related to participants' aspirations and goals for their language learning, whereas comments related to the suggestions related to participants' language learning experience and opportunities.

Among the three participants who consistently linked the report content with their own lives, reflections were additionally extended to the skill descriptions and figures. These comments focused on evaluating the test results and their own skills. Often, the comments indicated that the participants were thinking about their English language usage experience outside the classroom. For example:

Sometimes I found that if you give me a novel, a native [English] novel, and I can read all of them really quickly and I know every word's meaning but I just cannot understand 100 percently just like I was reading in Chinese.

Most of the time I have communication note for my turn, and I send all information for my son's teacher and she respond me we understand each other. Nowadays she prefer email and I send email for the information. We understand each other.

4.3.1.4 *Dealing with comprehension challenges*

Most of the interview participants encountered comprehension challenges, but their responses to these challenges differed. At one end of the spectrum, two participants expressed little to no comprehension challenges, although one of them was observed to miscomprehend the purpose of the introduction. Instead, these two participants' comments focused entirely on applying the information to their own lives (see above) or evaluating the information (see below).

A further three participants frequently reported meaning making processes as they read. They encountered comprehension challenges and reported seeking to understand and resolve the comprehension challenge, which in general they did. For example: "I was thinking, what's the word mean, because I didn't see that my opinion and the test opinion be compared these two part, so I was thinking what does it mean, *do I agree or not.*" and "I'm trying to get the information about *separating ideas*. What the *separating ideas* is. Uh (reading under breath). Yeah I just um, reading, so. Just scan it and, ok I got it. Yeah."

The other six participants reported comprehension difficulties that appeared unresolved. Sometimes they asked directly for resolution, and other times they simply acknowledged lack of comprehension. A particular point of initial confusion was the figures, with five participants

reporting unresolved comprehension difficulties. They were all concerned about the large gaps between the bars on the graphs, and wanted to understand what they meant. Comments were along the lines of “I think *have you mastered it*, I don't understand *tests says* and *you says*, I don't understand these.” It is possible that other than lack of vocabulary, another source of confusion was that nine of the participants were used to being told their English proficiency level in terms of Canadian Language Benchmark levels, and the absence of this framework confused their expectations of how to read the figures. However, only two of the participants reported seeking to match the figures to Canadian Language Benchmark levels, so participants did not in general appear to expect a Canadian Language Benchmark-matched feedback system.

Regarding other areas of the report, two participants commented on unresolved comprehension issues with the skill descriptions, for example:

Know the mood of a text. How should I know the mood of a text? How should I know the mood? I can know the topic, and know the... (pause). These questions, to me, are not that much self-explanatory, you know.

And another two commented on planning challenges, for example: “Ok so the part I have trouble with is, how to I evaluate myself? How does someone evaluate their own self???” (Translated from Chinese).

It is notable that despite the learning suggestions receiving a lot of attention from participants, no major comprehension problems were observed for this section of the report. In contrast, the suggestions prompted much critical evaluation, which is the final theme observed, discussed below.

4.3.1.5 Extent of critical evaluation of report content

The final topic observed during the eye-tracking recall interviews was the extent of critical evaluation in which participants engaged while reading their report. The majority of critical engagement observed took place while reading the suggestions for learning. However, four participants reported little to no critical evaluation of the report content. Their thoughts were focused on comprehending the report and, particularly in the suggestions for learning section, applying the content to their own lives. Three participants appeared to engage critically with the suggestions for learning in the sense that they evaluated the possibilities for applying them in

their own learning. However, note that they were not negatively critical, although some ambivalence was expressed. The following example represents the reported thoughts of these participants:

“Because I am used to looking for the how to use, how to learn English best way. And ah, the next sentence they show, yes, must improve your ability to use vocabulary. Yeah I think it’s a good thing, but uh, for me, I feel some difficult to remember. Remember. Because I, I want to remember some vocabulary today but next day I won't remember, maybe I forget it! So I will find how to keep remind me. Because in my memory when I learn some vocabulary in my young years I remember it clear but recently I find I'm more easy to forget everything! Yeah. So here's some, a plan to show how to read and how to improve, I think it's uh, more, more best for me, yeah. And ah... But I don't know how to use these cultural, use culture to improve the ability.”

Finally, four participants reported critical evaluation that included both positive and negative conclusions. Once again, this critical engagement was focused on the suggestions for three of the four participants. These participants evaluated the utility of the suggestions for themselves, and where negative conclusions were drawn, it was because they felt they already knew the suggestion, the suggestion was unrealistic, and/or it was too general. Examples of negative evaluations from these participants about the suggestions are:

I was wondering if they can give me right suggestion or not. And uh, I was thinking in maybe the suggestion could be wrong cause uh, they don't, they don't know what I really mean but they just using some tools and it's not gonna be correct (laugh)... And, it’s so general, uh yeah, I know that, everyone knows that, so. Nothing special, yeah, we all know that (laugh).

I was thinking, actually, to be honest, I was thinking, this one is useless because I will not just spend my time to guess and to see if I’m right. I will just read them all and guess some vocabularies and try to understand, but not- because this not useless, just I am too lazy to do it.

However, these participants also drew positive conclusions where they deemed the suggestions relevant and of suitable quality, for example “This one is more, like, practical, yeah...It tell you

how to do this, like, discuss with other people the information the writer wants to communicate... Yeah and it's very easy to do it. 'Coz it's not like, reading lots of information about similar topic.'

An additional note is that only two participants reported other types of critical engagement with other areas of the report – principally questioning the validity of the results. However, their engagement appeared quite different. For example, one participant appears to be engaging in meaning making: "I was thinking if the test is wrong, because I'm really not good at find out obvious meaning if it is not very clear. In contrast, the other participant is expressing helplessness and frustration: "It is not obvious. Sometimes the question was not 100% clear, ok."

Based on the variation observed within each of the identified topics, rich profiles emerged of each participants' interactions with the reports. However, no participant's profile was exactly the same as another's, reflecting both the small sample size and the complex paths that learners follow when interacting with feedback.

4.3.2 RQ 3.2: What processing outcomes do language learners report after receiving a feedback report?

In addition to observing in detail how participants interacted with their reports, the outcomes of these interactions and any subsequent cognitive processing were of interest because the thoughts and feelings that language learners take away from their reports will affect how they use their feedback. In order to assess participants' processing outcomes, all participants completed a survey after receiving their report that included Likert-style questions exploring the processing outcomes of interest. Three pairs of constructs were hypothesized and five constructs were subsequently observed in the responses. Each construct represented a desired outcome that, based on literature review, was believed to support use of feedback. The first construct sought to understand experienced cognitive load and perceived ability to process report content: sense of overwhelmedness. The second pair sought to understand participants' cognitive and affective outcomes with respect to the report: trust in the report content and reflecting on one's English skills. The third pair aimed to identify participants' intent to use the report for further learning: desire to use the report and feeling in need of help to use the report.

Descriptive analyses of mean response distributions for the five variables, depicted in Figure 17, indicated that many participants felt generally positive about the report. Four of the five constructs were very negatively skewed, with over fifty percent of participants scoring 4 or higher (on a scale of 1 of 5) on each construct. The one construct that was fairly normally distributed was sense of overwhelmedness, which had a mean of 3.03 and a standard deviation of 1.11. In contrast, the final construct: feeling in need of help to use the report, was skewed, with a population mean score of 3.87 ($SD = 1.13$) and 20% of participants holding mean scores of '5'.

The remaining three constructs, all of them positive outcomes for report processing, were extremely negatively skewed. Specifically, the population mean score for trust in report content was 4.08 ($SD = 0.93$) and 32% of participants scored a mean of '5'. The population mean score for reflecting on one's English skills was 4.40 ($SD = 0.78$) and 46% of participants scored a mean of '5'. And the most skewed variable of all was desire to use the report, with a population mean score of 4.54 and 55% of participants scoring a mean of '5'. Although the skewed nature of the distribution for these constructs creates analytical challenges, in terms of desired report processing outcomes, the number of strong scores observed on these constructs is positive.

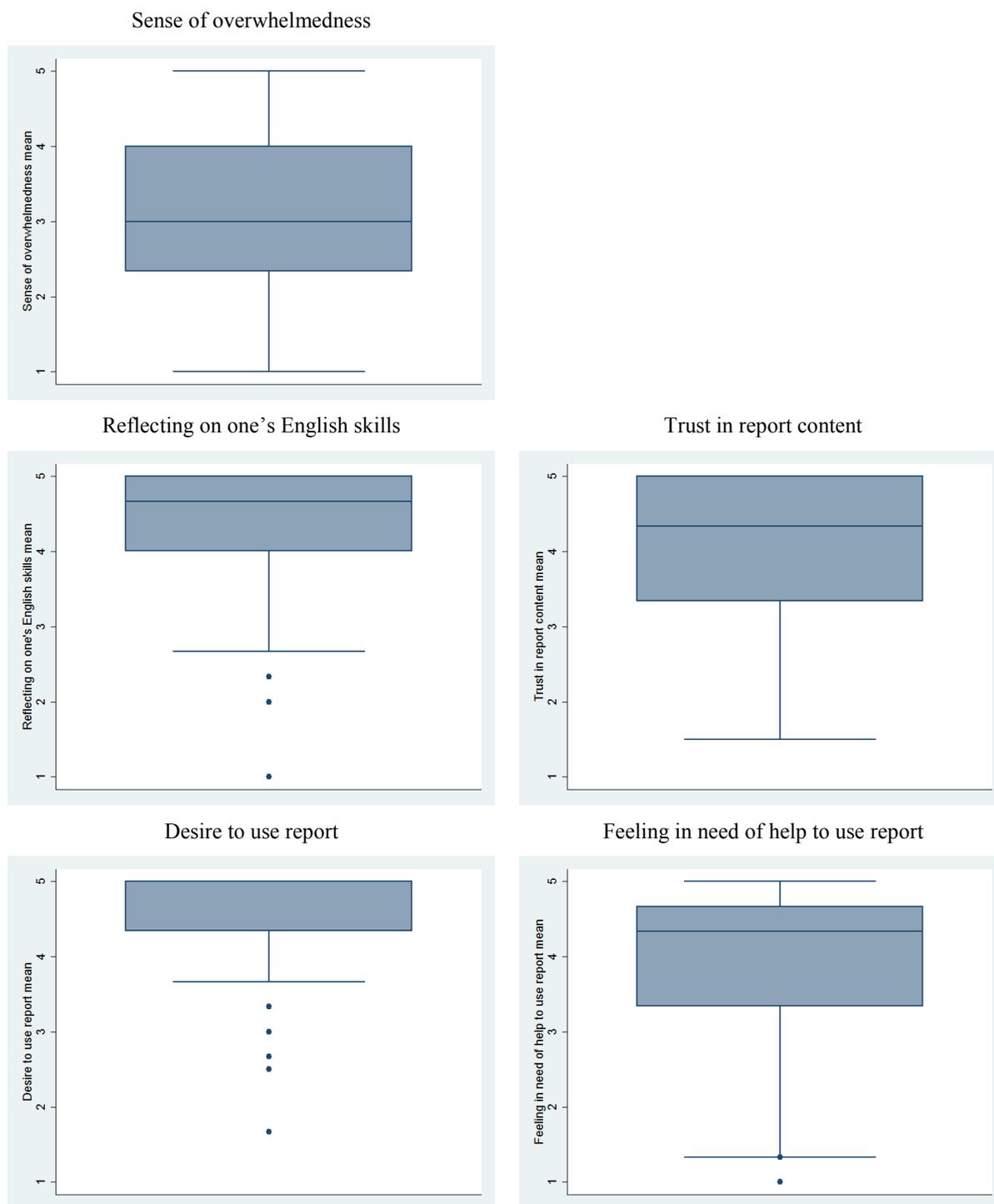


Figure 17. Box plots showing distribution of mean scores for each report processing construct. These overall trends appear to indicate that overall, participants were positive about the report; they generally trusted it, were reflecting on their own English skills, and wanted to use the report. Relatively fewer participants felt in strong need of help to use the report or paid attention to all

areas of the report, and participants were relatively unlikely to feel very overwhelmed. However, for all constructs there was a wide distribution of scores.

4.3.3 RQ 3.3: In what ways are language learners' processing outcomes related?

To investigate how processing outcomes related with each other, a latent profile analysis (LPA) was conducted to identify the ways in which participant profiles clustered. Two very similar solutions emerged as strong models. The overall best fitting comprehensive model was a four-class model that used the five mean composite variables identified in the survey data: sense of overwhelmedness, feeling in need of help, desire to use the report, trust in report content, and reflecting on English skills (adjusted BIC = 1191, loglikelihood value = -571, $N = 101$). Model fit is compared with other models in Table 34. The best fitting model also had entropy of 0.96, indicating strong classification quality. Note that class 3 is very small ($n = 4$), but attempting to omit the class in alternative models resulted in lower model fit, so it was retained.

Table 34

Participant processing: latent class profile models comparison table

Number of variables in model	Variable names	Number of classes	Adjusted BIC	Loglikelihood	Entropy	N
5	Overwhelmed-need help-want to use-trust-reflecting	2	1221	-594	.96	101
		3	1198	-578	.96	101
		4	1191	-571	.96	101
		5 **				
4	Overwhelmed-need help-want to use-reflecting	2	959	-467	.94	101
		3	932	-450	.96	101
		4 **				
		5 **				
3	Need help-want to use-reflecting	2	664	-324	.95	101
		3	645	-311	.94	101
		4 **				
2	Need help-want to use	3	436	-210	.94	100
		4	427	-203	.91	100
		5 **				

** Possible non-identification of model. Models with lowest group adjusted BIC in light grey highlight. The chosen model is in bold type and dark grey highlight.

Another model had better model fit (adjusted BIC = 427, loglikelihood value = -203, $N = 100$), but included only the two mean composite variables that addressed processing outcomes: feeling in need of help, and desire to use the report. Interestingly, when the means for all five of the variables included in the more comprehensive model were calculated, the same profiles emerged. The one exception is that one very small class was larger in the chosen model (that included only two variables), and this class' distinctive feature – a dramatically low mean for trust in the report content – became only moderately low. Overall, the two-variable model provided better statistical fit and its class profiles explained participant processing slightly better than the five-variable model. Therefore, the two-variable model that includes only processing outcomes is discussed below. Note also that there is a possibility that with a larger sample size, strong models that contain more variables could be identified.

Class 1, named in Figure 18 as 'no help, not use' contained 12% of participants, and these participants were overall least likely to want help using the report or want to use the report. The mean scores for feeling overwhelmed ($M = 2.64$, $SD = 1.03$) or in need of help ($M = 2.15$, $SD = 0.80$) were low for the population, similar to the 'no help, want to use' class (Class 4). However, their desire to use the report ($M = 3.07$, $SD = 0.60$), to be trusting the report content ($M = 3.53$, $SD = 1.09$), and to be reflecting on their English skills ($M = 3.56$, $SD = 1.21$) were also low for the population. However, note there was a wide range of scores for sense of overwhelmedness, trust in the report content and reflection on one's English skills, although means were consistently low. It would appear that the defining characteristics of Class 1 are little desire for help and weak desire to use the report.

Class 2, like the Class 1, contained 12% of the participants. Class 2 is labelled 'no help, want to use' on Figure 18. Like the Class 1, participants in Class 2 reported on average very low scores for overwhelmedness ($M = 2.41$, $SD = 1.31$) and feeling in need of help to use the report ($M = 2.21$, $SD = 0.66$), but in contrast to participant in Class 1, reported very high mean scores for the positive indicators of desiring to use the report ($M = 4.87$, $SD = 0.26$), trust in the report content ($M = 3.97$, $SD = 0.88$) and reflecting on one's English skills ($M = 4.64$, $SD = 0.50$). Moreover, the high mean for reflecting on one's English skills was consistent for the class, indicating a positive relationship for Class 2 membership with reflection.

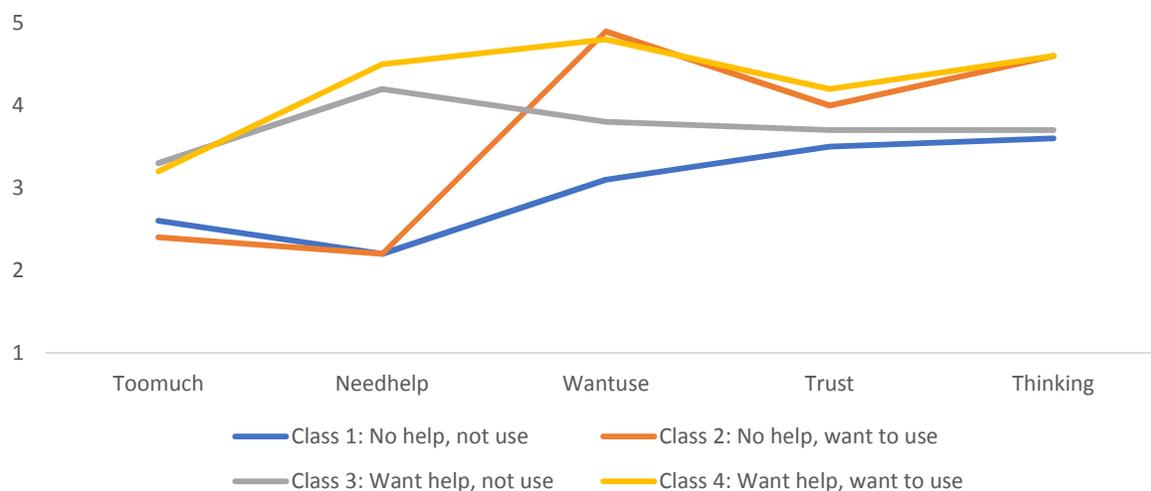


Figure 18. The mean scores on each variable, for each class

Class 3, containing 10% of the participants, is labelled ‘want help, not use’ on Figure 18. This class indicated a mixed profile, in which there was on average a relatively high sense of overwhelmedness ($M = 3.29$, $SD = 0.59$), only moderate interest in using the report ($M = 3.78$, $SD = 0.33$), and relatively little reported reflection on one’s English skills ($M = 3.67$, $SD = 0.78$). There was a relatively low sense of trust ($M = 3.72$, $SD = 0.95$) but also a high sense of needing help to use the report ($M = 4.20$, $SD = 0.45$). Mean scores score overwhelmedness were fairly consistently high, indicating a positive relationship for Class 3 membership with sense of overwhelmedness.

Class 4 was the largest class, labelled ‘want help, want to use’ on Figure 18, containing 66% of the participants. The means were consistently high for the population, with means of 4.84 ($SD = 0.25$) for desire to use the report, 4.24 ($SD = 0.88$) for trust in the report content, and 4.60 ($SD = 0.55$) for reflecting on one’s English skills. Participants in this Class 4 also believed they needed support, as the mean overwhelmedness score was high for the population ($M = 3.18$, $SD = 1.10$) and the mean score for feeling in need of help to use the report was the highest among the four classes ($M = 4.47$, $SD = 0.49$). Mean scores for reflection on one’s English skills were also fairly consistently high, indicating a positive relationship for Class 4 membership with reflecting on one’s English skills. Finally, the large size of this class is to be expected given the high skewedness of the variables, and indicates a large class of participants who wanted to use the report but feeling in need of help.

The hypothesis that the above variables represent the outcomes of positive interactions between the learners and their feedback is tentatively confirmed by the mean processing outcome scores for each of the interaction categories, discussed below. However, note that the smaller-than-expected sample size ($n = 11$) meant that statistical significance testing was not possible and findings should be interpreted alongside existing research.

Firstly, participants who minimized their personal responsibility for undesirable results ($n = 6$) scored a mean overwhelmedness score of 2.17 ($SD = 0.78$) compared to 3.27 ($SD = 0.76$) for those who did not minimize their personal responsibility ($n = 5$). Likewise, participants displaying emotional responses to their report ($n = 5$) had mean a overwhelmedness score of 3.27 ($SD = 0.86$) compared to 2.17 ($SD = 0.69$) for those who did not display emotional responses ($n = 6$). Relatively higher feelings of overwhelmedness ($M = 3.17$, $SD = 1.04$) were found in participants who did not report relating their report to the own lives ($n = 4$), while those who often related the report content to their own lives ($n = 3$) were generally not overwhelmed ($M = 2.00$, $SD = 1.00$).

Participants who often related the report to their own lives ($n = 3$) reported low mean feelings of needing help to use the report ($M = 3.11$, $SD = 1.90$), and similarly those who did not relate the report to their own lives ($n = 4$) had a mean 'in need of help' score of 3.50 ($SD = 0.58$). In contrast, participants who sometimes related the report to their own lives ($n = 4$) had a mean 'in need of help' score of 4.67 ($SD = 0.27$). Similarly, the participants who made no mention of needing help ($n = 2$) had a mean score of 2.00 ($SD = 1.41$) for feeling in need of help, compared to other participants ($n = 9$) who had a mean score of 4.22 ($SD = .62$) for feeling in need of help. Finally, participants with only positive conclusions from their critical engagement from the report ($n = 3$) had a mean score of 4.78 ($SD = 0.19$) for feeling in need of help, compared to 3.50 ($SD = 1.69$) for those who reported no critical engagement ($n = 4$) and 3.42 ($SD = 0.50$) for those with positive and negative conclusions from their engagement ($n = 4$).

Regarding desire to use the report, participants with emotional reactions ($n = 5$) had a mean score of 4.80 ($SD = 0.30$) compared to 4.22 ($SD = 1.36$) for those without emotional reactions ($n = 6$). Participants who often referenced their own lives ($n = 3$) were on average less likely to want to use the report ($M = 3.89$, $SD = 1.92$) compared to M s = 4.67 and 4.75 for other categories, $n = 8$, SD s = 0.67 and 0.32). Likewise, participants who gave no indication of comprehension

difficulties ($n = 2$) had a relatively low mean score for desire to use the report of 3.33 ($SD = 2.36$), compared to 4.72 ($SD = 0.53$) and 4.78 ($SD = 0.38$) for other categories ($n = 9$). Finally, participants with only positive critical evaluation of the report content ($n = 3$) all strongly wanted to use the report ($M = 5.00$, $SD = 0.00$), and those with positive and negative conclusions ($n = 4$) had a mean score of 4.83 ($SD = 0.33$), also high. This stood in contrast to those with no critical engagement ($n = 4$), who had a mean ‘desire to use report’ score of 3.75 ($SD = 1.50$).

Trust in report content and reflecting on one’s English skills both reflected the same relationships with report interactions. Participants who made no mention of comprehension challenges ($n = 2$) had relatively low mean trust and reflection scores ($Ms = 2.83$, $SDs = 0.71$ and 1.18) compared to other categories ($Ms = 4.30$ and 4.26 , $SDs = 0.61$ to 0.80 , $n = 9$). Likewise, participants who only reported positive conclusions from their critical engagement ($n = 3$) had higher mean scores for trust and reflection ($Ms = 4.56$ and 4.89 , $SDs = 0.77$ and 0.19 respectively) compared to other categories ($Ms = 3.83$ and 3.67 , $SDs = 0.82$ and 0.96 , $n = 8$).

The individual characteristics that learners bring with them to their report processing have been surveyed, an understanding of the attention that adult immigrant language learners pay to personalised feedback intended to support learning has now been established, and the types of interaction and processing outcomes that are observable have been described and profiled. The next research question turns to participants’ usage of the report, and first describes types of usage, then considers how attention, processing outcomes and background characteristics interact to affect learners’ report usage.

4.4 Theme 4 – Usage

4.4.1 RQ 4.1 How do language learners use a feedback report to plan their learning?

4.4.1.1 *Selecting skills*

In the planning section of their report, participants were required to select one skill, and could select up to three skills covered in the report, to work on in the coming weeks. Before reaching the planning section, participants were recommended to work on two specific skills in the suggestions section of the report, but they did not have to select those skills. Therefore, the first type of usage observable is the number and type of skills that participants selected to work on, and whether participants chose recommended skills.

Of the 100 participants who submitted a complete plan for at least one skill in the planning section, 15% selected only one skill, 24% selected two skills, and 61% selected three skills. In total, 70% of the participants chose to work on using vocabulary, making it the most commonly selected skill. The second most popular skill was inferencing, selected by 51% of the participants, and the third most popular was using cultural knowledge, selected by 39% of the participants. The remaining three skills were selected with approximately equal frequency; 30% of participants chose to work on using explicitly stated information, 30% chose to work on using implicitly stated information, and 26% chose to work on distinguishing ideas.

However, skill selection diverged substantially from the suggestions given to participants. Only 29% of the participants used both suggested skills in their planning, while 41% used one suggested skill. Fully 30% of participants used neither of the suggested skills. A key reason for this divergence is that the majority of participants were suggested to work on the skills of using culture and distinguishing information (55% and 59% respectively), but many selected to work on using vocabulary and inferencing. Specifically, 54% of participants did not receive a suggestion to work on vocabulary, but nevertheless chose to do so, and 40% of participants made the same choice for inferencing.

4.4.1.2 Setting goals

After participants selected a skill to work on, they were asked to write their learning goal for each skill – what they wanted to achieve. In all, 100 participants selected skills 248 times, and wrote out 196 goals. While the majority of participants who selected only one or two skills to work on all wrote associated goals (93% and 82% respectively), there was variation among participants selecting three skills to work on; only 62% of these participants wrote three goals.

Overall, 39% of all participants wrote three goals. However, the efficacy of the goals written varied substantially. Goals were coded according to whether they were skill-specific and/or observable. Skill-specific goals were clearly intended to practice the selected skill. Examples of skill-specific goals are: “to learn more new words to improve my speaking skill” (Using vocabulary), and “understand facts” (Using explicitly stated information), “To know the implied meaning of the sentence” (Using implicitly stated information), “Also i want write and talk very well, the connections is good for my english” (Inferencing), “I want to know what information is

important or relevant and what not” (Distinguishing ideas), and “getting to know more about Canadian Culture” (Using cultural knowledge).

In total, 40% of the participants made no skill-specific goals. Some goals were not skill-specific because they appeared to address other skills than the one selected, and many were not specific to any skill. For example: “separate ideas” (Using cultural knowledge), and “I want speaking better” (Using vocabulary).

Observable goals were distinct, concrete goals that a learner could focus on and monitor. Examples of observable goals are: “to learn 5 new words every day”, “At least, completely understand the direct stated information in articles”, “I can understand the information that not obvious in article”, “learn how to know what is coming next”, “understand reading separating idea clearly and realize and find topic quickly”, and “Understand idioms in Canada and phrase verbs”.

In total, 62% made no observable goals. Some of the ‘goals’ were not learning goals at all but planned actions, while many, as with non-skill-specific goals, were simply very general. For example: “I will go to library to get some reading text book. I will read them several times”, “To study in a university”, and “remembering more vocabulary so that easy to read”.

Generally speaking, goals first needed to be skill-specific, then they could also be task-specific; very few goals (6%) were task-specific and not also skill-specific. However, 48% of goals were neither skill-specific nor task-specific.

Notably, the skill-specificity and observability of goals varied by skill. ‘Using vocabulary’ elicited more skill-specific goals than would be expected by chance (68% of total) and elicited significantly more skill-specific goals than other skills, while ‘Using implicitly stated information’ and ‘Using explicitly stated information’ both elicited fewer skill-specific goals than would be expected by chance (33% and 21% respectively).

Certain skills also elicited more observable goals than others. Inferencing, using explicitly stated information, using vocabulary and using cultural knowledge all elicited fewer task-specific goals than would be expected by chance (33%, 21%, 20% and 11% respectively). Using cultural knowledge also elicited significantly fewer task-specific goals (11% of total) than using explicitly stated information (54%), distinguishing ideas (40%) and inferencing (33%), while

using implicitly stated information elicited significantly more task-specific goals (54% of total) than using explicitly stated information (21%), using vocabulary (20%) and using cultural knowledge (11%).

4.4.1.3 *Setting action plans*

In addition to selecting a goal for each skill selected, participants were invited to write an action plan for that skill. In all, 100 participants wrote out 204 action plans. While the majority of participants who selected only one or two skills to work on all wrote associated action plans (93% and 86% respectively), there was variation among participants selecting three skills to work on; only 60% of these participants wrote three action plans.

Several themes were observed in participants' action plans. In total, 32% of plans included study activities, and 25% used language from the report. In addition, 18% of action plans referenced the participant's life. Seven percent of action plans were repeated across multiple skills by the same participant.

As well as themes, the specificity of participants' action plans varied widely. Written plans ranged from not being a plan at all (e.g., "when I know this information I can direct some things to somebody", "I don't know") through to delineating specific activities (e.g., "everyday remember some vocabulary from newspaper or a book", "think more on the detail after reading and communicating"). Three main categories were identified, each of which was split into two or three subcategories.

The first category consisted of responses with limited evidence of planning, and included 'plans' that were not actually plans (see above, 6% of all action plans), plans that indicated a general intention to study, e.g., 'I will study hard', 'I will try a lot of practices in English learning' (12% of all plans), and plans to practice a modality (speaking, reading, writing, listening) with no further detail, e.g. 'maybe starting reading and writing every day', 'Speaking, more practice' (10% of all plans).

The second category included responses with moderate evidence of developing action plans. This category consisted of the 16% of action plans that specified a general activity to practice a modality, e.g., 'read more newspapers everyday and books, listen to radio and news', 'I thing read the books and the newspaper'. It also included the 22% of action plans that identified a

specific activity to practice a modality, e.g. ‘reading different books’, ‘Watch English tv series and pay attention to the subtitles trying to understand every conversation’.

Finally, the third category indicated specific activity plans. This category included the 19% of action plans that were general activities, e.g., ‘I will pay attention to my speech, and how I use new words in my speech’, ‘I will practice summarizing and sythesizing’. It also included the 15% of action plans that were specific activities, e.g. ‘ask others to communicate the main idea of text and ask their substitute expressions’, ‘Predicting what I will read next, then seeing if I am right, and thinking about and comparing what I know with what I read.’

As with goal setting, some skills elicited greater specificity in action plans than other skills. Inferencing, distinguishing ideas, and using cultural knowledge all elicited high proportions of specific action plans, with 45%, 47% and 41% of action plans for these skills falling into category 3 respectively. Using explicitly stated information and using implicitly stated information elicited high proportions of non-specific action plans; 42% and 41% of plans for these skills fell into category 1 respectively.

Skills also varied in whether report language was visibly used in the action plan. Only 16% of action plans for using vocabulary used language from the report. This was significantly fewer instances of report language than for inferencing, for which 34% of plans used report language, and for using cultural knowledge, 48% of the action plans for which used report language.

A notable theme in the choice of activities is that they were embedded in participants’ daily routines. For example, of 12 delayed recall interview participants, 10 wrote action plans that clearly reflected daily life. For example, to improve ability to use vocabulary: “I will help my children to study at schools-with their homework; I will continue taking my LINK classes; I will widen my communication with native speakers; I will continue any kind of reading, listening, watching and using.” And to improve ability to use cultural knowledge: “I listen every day to radio (news or talk show or discussions), and watch TV, daily shows which talk about daily events (like morning show). Participating in conversations and trips also help (like [at the local community centre]).”

4.4.1.4 *Setting monitoring plans*

Finally, for each skill that participants selected to work on, they were invited to write how they would monitor their progress. In total, the 100 participants wrote out 192 monitoring plans. Anecdotally, it can be noted that this was a challenging task for many participants, as they often asked how to fill out this section, but rarely asked how to fill out the other sections.

While all the participants who selected only one skill to work on wrote an associated monitoring plan, there was variation among participants selecting two or three skills to work on. Specifically, 73% of the participants who selected two skills also wrote two monitoring plans, and only 57% of the participants who selected three skills also wrote three monitoring plans.

Several themes were observed in participants' monitoring plans. In total, 18% of monitoring plans were repeated across multiple skills by the same participant, and 28% of monitoring plans were not actually monitoring plans, e.g. 'To participat at fastivals or events that hapend during the year', 'I hope good, because i like study english, i have voluntary'. Notably, 8% of monitoring plans mentioned using online activities. However, the major theme noted in participants' monitoring plans was the locus of control. Substantive monitoring plans coalesced into two groups: 26% of all monitoring plans had an internal locus requiring self-regulated monitoring of one's own performance, while 47% of monitoring plans had an external focus consisting of externally-regulated monitoring of performance, e.g., teacher, exams, exercises, judgment by proficient speakers. Examples of an internal locus of control are: 'I can read stories or news and guess what will be happend next and see is my gess right or wrong and why?', 'as far as i can speak with confidence i will notice', and 'If it is useful one word in many different situations'. Examples of an external locus of control are: 'speak with teacher', 'I can give test, then I know where I am', 'To practice with exercises on line', and 'ask people if they understand what i say'.

Note that many participants who wrote more than one monitoring plan used both internally and externally located monitoring plans. However, by the third monitoring plan, the percentage of 'non-plans' (see above) had risen to 38% of responses, compared to being only 24% of responses for the first monitoring plan. Moreover, a 'non-plan' was generally (but not always) the last monitoring plan written, indicating that the participants tended to use this strategy when they did not know what to write.

4.4.1.5 Planned discussion partners

Participants were most likely to want to talk to their teacher and/or friends and family, with the 100 participants reporting mean scores for 3.49 and 3.54 respectively, on a scale of 1 to 5, for probability of talking to each of these discussion partners about the report ($SDs = 1.58$ and 1.57). Specifically, 41% of participants were very likely to discuss their report with their teacher, and 43% of participants were very likely to discuss their report with their friends and family outside class. Participants were also likely to want to talk to their classmates, reporting a mean score of 3.21 ($SD = 1.57$); 30% of participants were very likely to discuss the report with these people. Notably, the distributions of responses for discussing the report with their teacher or classmates, shown in Figure 19, were u-shaped distributions. Twenty-one percent of participants were very unlikely to discuss their report with their teacher, and 25% of participants were very unlikely to discuss it with their classmates.

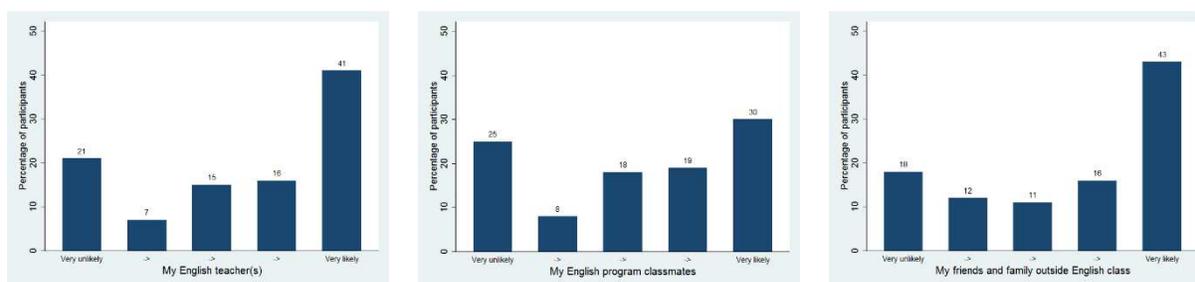


Figure 19. Distributions of which people want to talk to

4.4.1.6 Planned discussion topics

In terms of planned discussion topics, a hierarchy was observable regarding the frequency with which each topic was selected. The most commonly selected discussion topics were focused on improving English; 83% of participants wanted to discuss how to improve their English, and 74% want to discuss what areas of English skills to focus on. The next most-common topics were about understanding English skills; 64% of participants wanted to discuss what the report had said about their English language skills, 47% wanted to discuss their level of English proficiency, and 43% want to discuss how the self-assessment compared to their test results. Finally, the least commonly selected topics sought evaluation of the report's diagnosis; 26% wanted to discuss what the other person thought about their English skills, and 23% wanted to discuss how much they agreed or disagreed with the report.

Although participants were not asked who they wanted to discuss each of the topics with, Wilcoxon-Mann-Whitney tests were conducted to see whether participants selecting a specific topic were more likely to want to talk to certain people. These analyses provide a window on who they may be wanting to discuss these topics with. The tests indicated that participants who wanted to talk about how to improve their English were more likely to want to talk to their teacher than those who did not select this conversation topic ($Z = -2.40, p = .016$). Specifically, those participants wanting to discuss how to improve their English gave a mean score for likelihood of talking to their teacher of 3.67 (on a scale of 1 to 5, $SD = 1.49, n = 83$), compared to a mean of 2.59 ($SD = 1.73, n = 17$) for those who did not want to discuss this topic.

Likewise, the Wilcoxon-Mann-Whitney tests indicated that participants who wanted to talk about how to improve their English, what the other person thinks about their English skills, and what the report says about their English skills, were more likely to want to talk to their classmates ($M = 3.35, SD = 1.55, n = 83; M = 3.92, SD = 1.38, n = 26; \text{ and } M = 3.53, SD = 1.54, n = 64$ respectively) than those who did not select these conversation topics ($M = 2.53, SD = 1.50, n = 17, M = 2.96, SD = 1.56, n = 74, \text{ and } M = 2.64, SD = 1.46, n = 36$ respectively) ($Z = -2.02, p = .043; ; Z = -2.77, p = .006; Z = -2.75, p = .006$ respectively). Figure 20 reports the mean scores for wanting to talk with classmates about the report (on a scale of 1 to 5), for those participants who did and did not select each conversation topic.

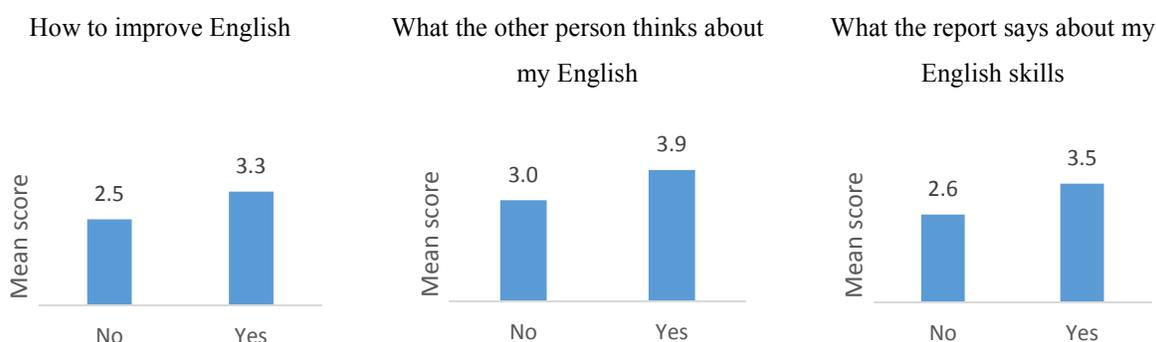


Figure 20. Mean 'talk to classmates' score by wanting to discuss specific topics

Finally, the Wilcoxon-Mann-Whitney tests failed to indicate any topics that identified participants who were more likely to want to talk to their friends and family outside class than those who did not select these conversation topics.

4.4.1.7 *Reported report usage after one month*

After one month, participants who were still willing and available, met and talked about what they recalled of the report, and whether they had used it. Among the recollections and reported usage for the 12 participants whose interviews were usable, several themes were identified. The most consistent theme was that everyone commented that they felt the report had said their English reading skills were lower than they expected. For example: “But I was very surprised because I don’t know which level represent the graphic but it’s very different from my level in my thinking!”, “I just remember, every skill...I said I can have a little bit, maybe I have some this kind of skill, but the machine said: no! Yeah, almost zero. Almost zero level, so I feel sad!”, and “Actually I remember that my expectations were more than my evaluation, my assessment! Even I minimize my expectations! I was surprise the result, honestly.”

Despite the consistent surprise about unexpectedly low results, participants dealt with the situation differently in terms of subsequent usage. The clearest variation in usage was that seven of the twelve delayed recall interview participants reported using the report and doing activities they wrote in their plans, while five did not. Of these five participants, three were positive about the report but cited external reasons for not doing planned activities, while two rejected the report.

A typical example of a participant feeling positive about the report but not using it was:

I didn’t use my report yet because I didn’t have time. I get sick last week so I didn’t, I didn’t use the report but I’m thinking about this, it’s important for the future. But I don’t know how to start. Because I’m trying since I study here or in another school, I start to read more, but I think it’s not enough. So maybe that is a reason for the report show me different things. But honestly, I didn’t use the report...Because this month I was very busy with my family, my mother in law visited me and I was shopping around and I don’t have time to focus in this.

It is a reality for immigrants with family care duties that utilizing reports and other study opportunities is sometimes just not possible outside dedicated program hours.

An example of a participant explaining rejection is:

So it's ok but it's with the timing and everything, I did ok but with the results see, it's little, I don't know about that. But I did ok in my view. Otherwise I'm ok with the reading. I understood everything, when they give us the [*unintelligible*] and the paragraph and the question, I understood. Sometime I don't have enough time, I think fourth section, no? One section, some of them I see, sometime really have more time, that's the problem. The time is the problem, otherwise it's not very hard. But I don't understand the report sometime, you know.

It is likely that there are individual reasons why each of the two participants who rejected the report did so. However, it should be noted that both participants belonged to the same class, and the other participant from that class also did not use the report, although was positively inclined toward it. This class was also the only class participating in the delayed recall interviews where no one used the report. This observation is notable because the class was also the only class where the teacher and student body openly and collectively expressed dissatisfaction with the report, stating that they thought the reported achievement was too low. These findings are an indication of the importance of the social context in which the report, or any feedback is received; content will be rejected if it is presented in a skeptical or hostile context.

A difference in written plans was also observable among these usage groups. Of the seven participants who reported doing planned activities, six participants' written plans contained explicit reference to either the skill descriptions or suggestions that they saw on their report. In contrast, only two of the five participants who reported not doing any planned activities originally used language from their report in their plans. Both of these participants were positive about the report; the two who rejected the report had not used report content in their written plans and in fact did not write any substantive plans at all. It would appear that rejection of the report starts soon after receiving it.

Among the seven participants who reported doing at least some of the activities they wrote in their plans, a key observation is that they all co-opted the report into their own schemes of learning. That is to say, rather than follow the report's recommendations directly, participants selectively used skill descriptions and suggestions in their own plans. Specifically, the participants had their own ideas about what would help them and about what they had time and

resources to do, and they made plans that fit into this personal learning scheme. The report content was utilized to the extent that it fit within these schemes. This was previously observed in the tendency for lower proficiency participants to select to improve their vocabulary regardless of report suggestions. Another outcome of this approach was that suggestions that emphasized daily life rather than formal study were written and done much more frequently, as shown in Table 35. In particular, reading newspapers, listening to the radio and watching TV were common report content that were included in plans. As noted previously, these types of daily activities were also popular among all participants' plans.

Table 35

A comparison of report content that was observably used in delayed recall interview participants' written plans, and how it was used in participants' written plans

Example	Report content	Participant written plans content
A	listening to radio news...	Listen to Radio
B	volunteering with local organizations and at events	To make some volunteering
C	understand informal words and phrases	understand informal words and phrases and use
D	listening to radio news...	I listen every day to radio (news or talk show or discussions)
E	understand information that is not obvious	My goal is to understand the information that are not obvious in the text...
F	watching movies and TV with subtitles	I will use the vocabulary, watching TV... I will watch movies
G	find useful information and ignore unimportant information... can see the main ideas and the details	I want to know what information is important or relevant and what not... Find the main idea and detect the details that help me to understand.
H	listening to radio news and reading newspapers daily... taking part in Canadian holidays, festivals and hobbies	I will read the newspaper. I will listen to radio news and TV news. I will try to read some books about Canadian culture and history of Canada. I will try to read about traditional ceremonies.

Finally, five of the seven participants who did the activities in their plans also mentioned doing additional activities based on the report content, that they did not write in the Planning section of their reports, listed in Table 36. This finding indicates that participants were paying attention to more report content than they wrote in their plans.

Table 36

Additional activities reported by delayed recall interview participants that were also mentioned in their report

Additional activities
listening the CBC radio
I saw the report, now I try to catch the main idea when I write or speaking, like that
I try to listen CBC news
You recommend me to listen more and reading the newspaper for improving my knowledge about the Canadian cultures. And customs. And try to do this one month ago... I write the vocabulary, I find it in dictionary and try to remember it.
I have borrowed some story books from the library and I'm going to start this week to reading these book one hour a day and after I finished a book I'm going to summarize the book in order to practice my writing and reading, and I can check if I learn many words and if I can understand the story, I can write it down

4.4.2 RQ 4.2 What are the relationships between individual characteristics, attention and processing experience, and usage of feedback?

A major question for report usage is what factors impact usage. To answer this question, inferential statistical analyses were conducted, including correlational analyses, a Kruskal-Wallis test, and linear, ordinal and logistic regression analyses. The analyses determined the relationship of observed psychological characteristics, reported attention and reported processing experience with usage of the report planning section, and with intended discussion interlocutors and topics.

4.4.2.1 *Written plans: Influences on number of skills selected and choosing more suggested skills*

First, the relationships between factors (including psychological characteristics, reported attention and reported processing experience) and selecting more skills to work on, and selecting more suggested skills to work on, were investigated. In terms of correlations, number of skills selected (1 to 3) was positively and significantly associated with number of report sections paid a lot of attention to ($\rho = .30, N = 97, p = .003$), and extent of attention to skill descriptions ($\rho = .39, N = 97, p < .001$), graphs ($\rho = .23, N = 94, p = .025$), suggestions for learning ($\rho = .23, N = 95, p = .023$), and planning ($\rho = .23, N = 95, p = .026$).

Meanwhile, number of suggested skills selected (0 to 2) was positively and significantly associated with raw total score ($\rho = .49, N = 100, p < .001$) and self-assessment score ($\rho = .32, N = 100, p = .004$), attention to skill descriptions ($\rho = .28, N = 97, p = .002$) and graphs ($\rho = .27, N$

= 94, $p = .007$), and total number of skills selected ($\rho = .48$, $N = 100$, $p < .001$). Number of suggested skills selected was significantly negatively associated with feeling in need of help to use report content ($\rho = -.24$, $N = 99$, $p = .004$).

Therefore, the only variables associated with number of skills selected during planning were variables indicating extent of reported attention to report sections. A Kruskal-Wallis test on number of sections paid much attention to (0 to 5), with number of skills selected (1, 2 or 3) as the independent variable indicated that participants who reported paying attention to several parts were significantly more likely to select more skills in the planning section ($\chi^2(2) 8.56$, $p = .014$). Among participants who reported spending a lot of time or all their time to four or five sections of the report, 79% selected three skills. In comparison, 66% of those who reported spending a lot of attention to two or three sections choose three skills, while only 45% of those who paid a lot of attention to only one section or no sections selected three skills. This decrease is shown in Figure 21.

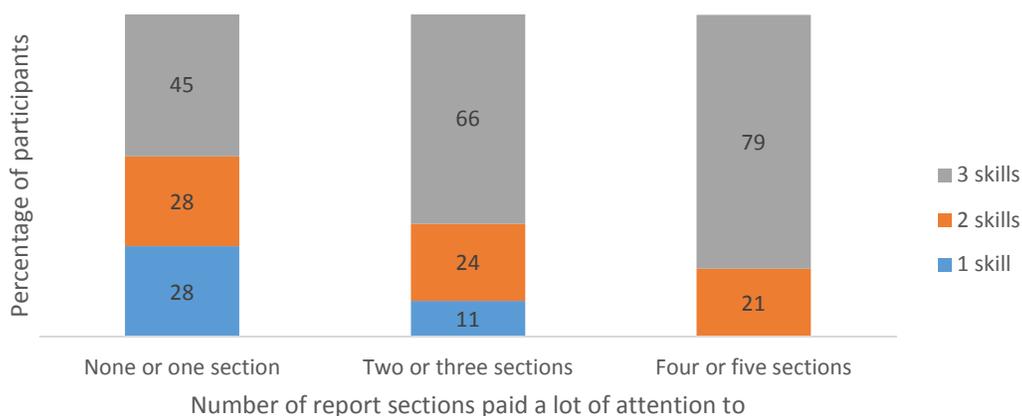


Figure 21. Percentage of participants choosing each number of skills, by number of report sections paid a lot of attention to

Regarding whether attention to any specific report sections was associated with selecting more skills to work on, an ordinal regression indicated that, taking into account shared variances, an increase in reported attention to both skill descriptions and suggestions – there was an interaction effect – was associated with an increase in the odds of selecting more skills, with an odds ratio of 1.19 (95% CI, 1.08 to 1.31), likelihood ratio $\chi^2(1) = 14.88$, $p < .001$. For example, based on the model's regression equation, participants who reported a lot of attention to both skills

descriptions and suggestions had a .79 probability of selecting three skills. Those reporting low attention to both report sections had only a .31 probability of selecting three skills.

However, choosing more skills to work on is not necessarily the same as choosing more suggested skills. An ordinal regression indicated that taking into account shared variances, choosing more of the recommended skills was significantly associated only with raw total test score (a proxy for English reading proficiency), and the overall number of skills chosen. Specifically, an increase in total test score and in number of skills selected was associated with an increase in the odds of selecting more suggested skills, with an odds ratio of 1.16 (95% CI, 1.09 to 1.23) for total test score and an odds ratio of 4.71 for number of skills selected, likelihood ratio $\chi^2(2) = 54.65, p < .001$. For example, based on the model's regression equation, participants who selected three skills and scored 10 out of 38 had a .15 probability of selecting both suggested skills, while participants who selected three skills and scored 30 out of 28 had a .78 probability of selecting both suggested skills.

4.4.2.2 *Written plans: How participants set goals*

Next, the relationships between background factors and writing more goals overall, writing more skill-specific goals, and writing more task-specific goals, were investigated. An interesting hierarchy of predictive relationships emerged between number of goals written, number of skill-specific goals, and number of observable goals.

First, an ordinal regression analysis indicated that the odds of writing more goals were associated with increases in the number of skills selected, the number of skill-specific goals written, and attention to skill descriptions, with an odds ratio of 6.77 (95% CI, 2.85 to 16.07) for number of skills selected, an odds ratio of 2.14 (95% CI, 1.19 to 3.84) for number of skill-specific goals written, and an odds ratio of 1.89 (95% CI, 1.05 to 3.40) for amount of attention to skill descriptions. Higher language proficiency was associated with a decrease in the odds of writing more goals, with an odds ratio of 0.88 (95% CI, 0.81 to 0.94). The model had likelihood ratio $\chi^2(4) = 60.69, p < .001$. For example, based on the model's regression equation, a participant selecting three skills who scored 10 out of 38 had a .75 probability of writing three goals, while a participant who selected three skills and scored 30 out of 38 had a .11 probability of writing three goals.

Second, another ordinal regression analysis indicated that the odds of writing more skill-specific goals were associated with increases in number of goals written and number of observable goals written, with an odds ratio of 2.37 (95% CI, 1.43 to 3.95) for number of goals written, and an odds ratio of 6.10 (95% CI, 3.18 to 11.70) for number of observable goals written. The model had likelihood ratio $\chi^2 (2) = 56.38, p < .001$. For example, based on the model's regression equation, a participant who wrote three goals but wrote no observable goals had a .13 probability of writing two skill-specific goals, while a participant who wrote three goals and one observable goal had a .44 probability of writing two skill-specific goals.

Finally, a third ordinal regression analysis showed that the odds of writing more observable goals were only associated with an increase in number of skill-specific goals written, with an odds ratio of 6.18 (95% CI, 3.30 to 11.56), likelihood ratio $\chi^2 (1) = 45.33, p < .001$. For example, based on the model's regression equation, a participant who wrote no skill-specific goals had a .08 probability of writing one observable goal, while a participant who wrote two skill-specific goals had .46 probability of writing one observable goal.

4.4.2.3 Written plans: How participants set action plans

As well as goal setting behavior, the relationships between background factors and writing more action plans, and specificity of plans, were investigated. For the number of action plans written, an ordinal regression indicated that selecting more skills to work on was, logically, associated with an increase in the odds of writing more action plans, with an odds ratio of 8.61 (95% CI, 4.03 to 18.39). Additionally, a stronger performance avoid goal orientation and a higher raw total test score were both associated with a decrease in the odds of writing more action plans, with odds ratios of 0.49 (95% CI, 0.27 to 0.87) and 0.90 (95% CI, 0.84 to 0.96) respectively. The model had likelihood ratio $\chi^2 (3) = 53.47, p < .001$. For example, based on the model's regression equation, a participant who scored 10 out of 38 and held a weak performance avoid orientation had a .84 probability of writing three action plans, compared to a probability of .66 for a participant who scored 10 out of 38 and held a strong performance avoid goal orientation, a probability of .40 for a participant who scored 30 out of 38 and held a weak performance avoid orientation, and a probability of just .20 for a participant who scored 30 out of 28 and held a strong performance avoid orientation.

In terms of plan specificity (an ordinal variable on a 0 to 12 scale with an acceptably normal distribution), a linear regression analysis indicated that after controlling for the number of action plans written, plan specificity was associated with more reflection on one's English skills (standardized $\beta = .15$, $t(96) = 2.39$, $p = .019$), more attention to the skill descriptions (standardized $\beta = .14$, $t(96) = 2.25$, $p = .027$), and a lower performance prove goal orientation (standardized $\beta = -.14$, $t(96) = -2.34$, $p = .021$). This model explained 65% of variance in action plan specificity, $R^2 = .65$, $F(4, 91) = 45.28$, $p < .001$. For example, based on the model's regression equation, a participant who wrote three action plans, was thinking about their English skills and held a low performance prove goal orientation, was predicted an action plan specificity score of 9.1 (out of 12). In contrast, a participant who also wrote three action plans but was not thinking much about their English skills and held a high performance prove goal orientation was predicted an action plan specificity score of 6.5.

4.4.2.4 *Written plans: How participants set monitoring plans*

The final investigation of participants' use of the planning opportunity concerned how they set monitoring plans. As previously discussed, monitoring plans were observed to group into internally regulated and externally regulated methods. Therefore, as well as using an ordinal regression model to identify factors associated with writing more monitoring plans, logistic regressions were conducted to identify the factors associated with writing at least one internally regulated, at least one externally regulated monitoring plan, and only writing monitoring plans that showed monitoring (either internal or external).

Regarding number of monitoring plans written, an ordinal regress analysis showed that after controlling for number of skills selected, the odds of writing more monitoring plans were associated with lower raw total test score (odds ratio 0.90, 95% CI 0.84 to 0.96), higher trust in the report content (odds ratio 1.97, 95% CI 1.14 to 3.39), and repeating monitoring plans across skills (odds ratio 16.42, 95% CI 3.48 to 77.45). The model had likelihood ratio $\chi^2(4) = 63.84$, $p < .001$. For example, based on the model's regression equation, for participants who selected three skills and did not repeat monitoring plans, those who scored 10 out of 38 and had high trust in report content had .77 probability of writing three monitoring goals. Those who scored 30 out of 28 and had low trust in the report content had just .11 probability of writing three monitoring goals.

Moreover, a logistic regression analysis indicated that once number of monitoring plans was controlled for, the odds of writing at least one internally regulated monitoring plan were associated with not repeating monitoring plans (odds ratio 0.04, 95% CI 0.00 to 0.38), and weaker fixed beliefs about intelligence (odds ratio 0.61, 95% CI 0.39 to 0.95). For example, based on the model's regression equation, among participants who wrote three monitoring plans, a participant with weak fixed beliefs about intelligence who did not repeat their monitoring plans had a .82 probability of writing at least one internally regulated monitoring plan. Meanwhile, a participant who wrote three monitoring plans, held strong fixed beliefs about intelligence and did not repeat monitoring plans had a .64 probability of writing at least one internally regulated monitoring plan, and similar participant who repeated some monitoring plans had a .06 probability of writing at least one internally regulated monitoring plan.

In contrast, another logistic regression analysis showed that the odds of writing at least one externally regulated monitoring plan was not significantly associated with the number of monitoring plans written. The factors associated with increased odds of writing at least one externally regulated monitoring plan were higher total raw test score, with an odds ratio of 1.06 (95% CI, 1.00 to 1.13), and a stronger mastery goal orientation, with an odds ratio of 2.43 (95% CI, 1.09 to 5.41). The model had likelihood ratio $\chi^2(3) = 11.45, p = .009$. For example, based on the model's regression equation, a participant who wrote three monitoring plans, scored 30 out of 38, and held a strong mastery goal orientation had a .89 probability of writing at least one externally regulated monitoring plan. A similar participant who score 10 out of 28 and held a weak mastery goal orientation had a .51 probability of writing at least one externally regulated monitoring plan.

Finally, a third logistic regression analysis showed that the odds of writing at least one monitoring plan that showed monitoring (either internal or external) were associated with writing fewer monitoring plans (odds ratio 0.36, 95% CI 0.18 to 0.69), weaker fixed beliefs about intelligence (odds ratio 0.59, 95% CI 0.34 to 0.91), less sense of overwhelmedness (odds ratio 0.57, 95% CI 0.34 to 0.97), and a stronger mastery goal orientation (odds ratio 6.69, 95% CI 2.32 to 19.28). The model had likelihood ratio $\chi^2(4) = 31.01, p < .001$. For example, based on the model's regression equation, a participant who wrote three monitoring plans, held weak fixed beliefs about intelligence and a strong mastery goal orientation, and was not feeling

overwhelmed, had a .81 probability of writing at least one monitoring plan that showed monitoring of some kind. In contrast, a participant who wrote three monitoring plans, held strong fixed beliefs about intelligence, a relatively weak mastery goal orientation, and was feeling quite overwhelmed had a .10 probability of writing at least one monitoring plan that showed monitoring of some kind.

4.4.2.5 *Planned discussions: Who learners want to talk to*

In addition to observing how participants used the planning section of the report, participants were also asked who they were likely to talk to about the report, and what they wanted to talk about. No significant associations were observable in ordinal regression models for wanting to talk to classmates or friends and family, although lower test scores were significantly associated with a stronger desire to speak to classmates ($\rho = -.23, N = 100, p = .022$), and a stronger desire to speak to friends and family was significantly associated with a greater desire to use the report ($\rho = -.23, N = 100, p = .045$), and having paid more selective attention to report content ($\rho = -.24, N = 99, p = .016$).

However a third ordinal regression showed that the odds of a participant reporting that they were very likely to talk to their teacher increased as total raw test score decreased (odds ratio 0.91, 95% CI 0.87 to 0.96), and as mastery goal orientation increased (odds ratio 2.17, 95% CI 1.10 to 4.26). The model had likelihood ratio $\chi^2(2) = 17.78, p < .001$. For example, based on the model's regression equation, a participant who scored 10 out of 38 and held a high mastery goal orientation had a .66 probability of being very likely to want to talk to their teacher about the report. Meanwhile, a participant who scored 10 out of 38 and held a relatively low mastery goal orientation had a .50 probability of being very likely to want to talk to their teacher, and a participant who scored 30 out of 38 and held a relatively low mastery goal orientation had a .13 probability of being very likely to want to talk to their teacher about the report.

4.4.2.6 *Planned discussions: What learners want to discuss*

After participants were asked who they were likely to talk to about the report, they were asked what they would probably talk about. The results of logistic regressions on each of the topics were very interesting, and will be addressed in order of the frequency by which discussion topics were selected.

The most commonly selected topic, how to improve my English, was selected by 83% of participants. The odds of selecting this topic decreased as test score increased (odds ratio 0.93, 95% CI 0.87 to 1.00) and performance prove goal orientation increased (odds ratio 0.33, 95% CI 0.15 to 0.73). The odds of selecting this topic increased as feelings of needing help to use the report increased (odds ratio 1.85, 95% CI 1.12 to 3.06). The model had likelihood ratio $\chi^2(3) = 16.42, p < .001$. For example, based on the model's regression equation, a participant who scored 10 out of 38, had a weak performance avoid orientation and felt strongly in need of help to use the report had a .98 probability of wanting to discuss how to improve their English. A participant who scored 30 out of 38, held a relatively strong performance prove goal orientation and did not feel in great need of help to use the report had a .54 probability of wanting to discuss how to improve their English.

The next commonly selected topic, what areas of my English skills I should focus on, was selected by 74% of participants. The odds of selecting this topic decreased as sense of overwhelmedness increased, with an odds ratio of 0.59 (95% CI, 0.36 to 0.98). The odds of selecting this topic increased as attention to figures increased (odds ratio 2.25, 95% CI 1.21 to 4.18), trust in the report content increased (odds ratio 1.98, 95% CI 1.12 to 3.53), and as feeling in need of help to use the report increased (odds ratio 1.66, 95% CI 1.02 to 2.72). The model had likelihood ratio $\chi^2(5) = 16.69, p = .005$. For example, based on the model's regression equation, a participant who held a low performance prove goal orientation, paid a lot of attention to the figures, was not overwhelmed, trusted the report content, and felt in need of help to use the report, had a .98 probability of wanting to discuss what areas of their English skills they should focus on. In contrast, a participant who held a relatively high performance prove goal orientation, did not pay much attention to the figures, felt overwhelmed, had relatively low trust in the report content and relatively low feelings of needing help to use the report, had a .19 probability of wanting to discuss what areas of their English skills they should focus on.

Next, the topic of what the report says about my English language skills was selected by 64% of participants. The odds of selecting this topic increased as self-assessment score increased (odds ratio 2.44, 95% CI 1.15 to 5.18), and as fixed beliefs about intelligence increased (odds ratio 1.70, 95% CI 1.15 to 2.52). The model had likelihood ratio $\chi^2(2) = 11.96, p = .003$. For example, based on the model's regression equation, a participant with a relatively high self-assessment

score and strong fixed beliefs about intelligence had a .83 probability of wanting to discuss what the report said about their English skills. A participant with a relatively low self-assessment score and weak fixed beliefs about intelligence had a .46 probability of wanting to discuss this topic.

Fourth, 47% of participants intended to discuss their level of English proficiency. The odds of selecting this topic increased as self-assessment score increased (odds ratio 2.92, 95% CI 1.38 to 6.21), and as feeling in need of help to use the report increased (odds ratio 1.55, 95% CI 1.02 to 2.36). The model had a likelihood ratio $\chi^2(2) = 10.60, p = .005$. For example, based on the model's regression equation, a participant with a relatively high self-assessment and strongly feeling in need of help to use the report had a .65 probability of wanting to discuss their level of English proficiency. Conversely, a participant with a relatively low self-assessment score and not particularly feeling in need of help to use the report had a .32 probability of selecting this topic.

Fifth, 43% of participants wanted to discuss how their self-assessment compared to their test results. The odds of selecting this topic increased as reflection on one's English skills increased (odds ratio 3.08, 95% CI 1.44 to 6.59), attention to skill descriptions increased (odds ratio 1.86, 95% CI 1.14 to 3.03), and performance prove goal orientations increased (odds ratio 1.76, 95% CI 1.03 to 3.01). This model had a likelihood ratio $\chi^2(3) = 19.06, p < .001$. For example, based on the model's regression equation, a participant with a relatively strong performance prove goal orientation, who paid a lot of attention to the skill descriptions and was reflecting a lot on their English skills, had a .77 probability of wanting to discuss how their self-assessment compared to their test results. A participant with a weak performance prove goal orientation, who paid little attention to the skill descriptions and was not really reflecting on their English skills, had a .02 probability of selecting this discussion topic.

A relatively rare desired discussion topic was what the other person thought about the participant's English skills; 26% of participants wanted to discuss this topic. The odds of selecting this topic increased as performance prove goal orientation increased (odds ratio 1.87, 95% CI 1.05 to 3.33), and as trust in the report content decreased (odds ratio 0.60, 95% CI 0.37 to 0.97). This model had a likelihood ratio $\chi^2(2) = 9.12, p = .010$. For example, based on the model's regression equation, a participant holding a relatively strong performance prove goal orientation but who was relatively untrusting of the report content had a .40 probability of

planning to discuss what the other person thought about their English skills. A participant with a low performance prove goal orientation and strong trust in the report content had a .11 probability of selecting this topic.

Finally, 23% of participants wanted to discuss how much they agreed or disagreed with the report. The odds of selecting this topic increased as mastery goal orientations increased (odds ratio 3.66, CI 1.23 to 10.91), and as desire to use the report decreased (odds ratio 0.49, 95% CI 0.26 to 0.93). This model had a likelihood ratio $\chi^2(2) = 9.98, p = .007$. For example, based on the model's regression equation, a participant holding a strong mastery goal orientation and but only moderate desire to use the report had a .34 probability of wishing to discuss how much they agreed or disagreed with the report. Conversely, a participant with a relatively weak mastery goal orientation but strongly wanting to use the report had a .10 probability of selecting this topic.

Chapter 5: Discussion

The objectives of this study were to investigate how English language learners engage with and process feedback differently, to better understand the relative importance of different learner characteristics, and cognitive and affective processing strategies. The research questions were organized into themes, and were:

- Theme 1 – Characteristics of adult immigrant English language learners in Canada
 - What are their occupational contexts?
 - What are their English language environments?
 - What are their English language goals?
 - What are typical goal orientation profiles and beliefs about intelligence?
 - What are typical English reading proficiency profiles?
 - How do they perceive their English reading proficiency?
 - What relationships are observable between individual characteristics?
- Theme 2 – Relationship between attention to and processing of feedback
 - To which aspects of feedback reports do language learners report paying attention?
 - To which aspects of feedback reports are language learners observed to pay attention?
 - Which aspects of feedback reports can be recalled by language learners one month later?
- Theme 3 – Language learners’ affective and cognitive interaction with report content
 - What affective and cognitive strategies do language learners report when receiving a feedback report?
 - What affective and cognitive processing outcomes do language learners report after receiving a feedback report?
 - In what ways are language learners’ processing outcomes related?
- Theme 4 – Usage of feedback
 - How do language learners use a feedback report to plan their learning?
 - What are the relationships between individual characteristics, attention, processing outcomes, and usage of feedback?

5.1 Discussion 1: Relationship between reported and observed attention to feedback

Based on the findings described in the results for Theme 2, reported and observed attention diverged substantially, and this divergence is summarized in Figure 22.

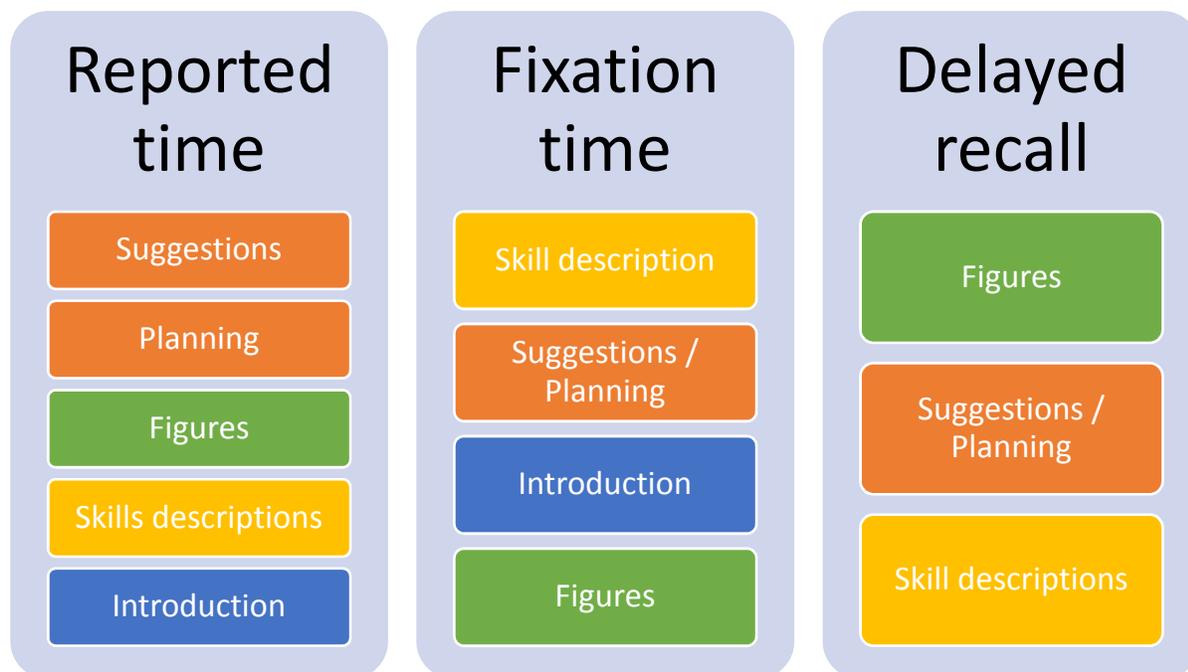


Figure 22. Relative amount of attention noted in the relevant data sources: Sections with most attention listed first

When asked which areas they spent the most time on, participants reported spending the most time on suggestions and planning ($M = 3.61$ and 3.42 , $SD = 0.93$ and 1.07 respectively, on a scale of 1 to 5), followed by figures ($M = 3.20$, $SD = 0.99$). The observed time spent on each report section contradicted these claims, as the most time was clearly spent on skill descriptions (28% of time on average), closely followed by suggestions and planning (25% and 24% of time on average respectively). In the observed data, it was the figures that participants spent very little time looking at (5% of time on average).

This contradiction can be explained from the eye-tracking interview data. The eye-tracking interview data indicated that the amount of time reported on each report section was closely related to the amount and depth of cognition undertaken by participants for that section.

Therefore, it is reasonable to conclude that when asked to report time spent, participants in fact reported amount of cognitive attention.

However, the third data set concerning attention – the delayed recall interviews – added further information about the figures. Although participants reported a moderate amount of attention to the figures – 37% reported spending ‘a lot of time’ or ‘all their time’ on the figures – and were observed to spend very little time looking at the figures, the figures and the information in them were clearly the most memorable information on the report, recalled by 13 of 14 delayed recall interview participants (93%). In contrast, suggestions for learning and planning were recalled by a similar proportion (50% and 57% respectively) of participants who reported spending ‘a lot of time’ or ‘all their time’ on these sections (58% and 52% respectively). Likewise, only three participants (27%) recalled skill descriptions, and a similar percentage (25%) reported spending ‘a lot of time’ or ‘all their time’ on these sections. Note also that the eye-tracking trace data explains this discrepancy for figures by showing how participants used the figures. Although little time was spent looking at them, they were a ‘framework’ for understanding the other information in the report, with users switching back and forth – if the figures had useful information – between figures and other available information.

Therefore, the relationships between reported time spent looking at each section, eye-tracking interviews, and the delayed recall interviews appear to indicate that adult language learners can report attention fairly reliably. Note that in contrast, the amount of time spent on a section is not directly related to attention; text-heavy sections simply require more time to read. In fact, there was an inverse relationship between amount of text in a section and the chances of that section being recalled one month later. However, the amount of text was also very much related to extent of personalization; the un-personalized sections were very unlikely to be recalled. Moreover, there was a skill profile effect; lower English reading proficiency participants had flat skill profiles and therefore paid little attention to the figures due to absence of useful information.

In conclusion, it appears that participants paid most attention to information that helped their learning (suggestions and planning) at the time of receiving the report and did indeed often recall these sections subsequently. However, the section with the most powerful long-term recall was the figures, which had been used as a frame of reference when interpreting other information in the report such as the suggestions for learning and planning, and thus left a strong impression. In terms of observed and reported attention, this study’s findings cohere with existing knowledge of visual processing and attention (Kang, 2014; Bisson, van Heuven, Conklin, & Tunney, 2014;

Owen, 2016; Wadlinger & Isaacowitz, 2011), and have implications for design of language learning feedback and of usage of various stimulated recall methods.

Regarding why participants did not accurately report the actual time that they spent looking at each section, research points to the automaticity of many decisions about directing attention (Uusberg, Uibo, Kreegipuu, & Tamm, 2013). First, the eye-tracking data clearly reflects levels of cognitive demand and processing weight, with more demanding report sections taking more time to read. This finding confirms existing research in cognitive load (cf. Ayres & Paas, 2012), which makes the claim that the ability to process information is affected by the relative load that the individual experiences at that time (Sweller, Ayres, & Kalyuga, 2011). A primary source of cognitive load is unfamiliar text, particularly when compared to figures, and particularly for language learners (Fontanini & Braga Tomitch, 2009; Segalowitz & Frenkiel-Fishman, 2005). Indeed, Kang's (2014) work in reading strategies likewise notes that second language learners read more slowly than first language readers.

However, the self-reported time spent on each area of the report shortly after receiving the report appears to primarily reflect participants' actual attention rather than actual time spent looking at each section. At this point the well-known differences between attention and gaze (Richardson, Dale, & Spivey, 2006; Posner, 1980; Rayner, 1998) become apparent, and the automaticity effect comes into play (Uusberg et al., 2013). The principle aspect to note is the relationship of affect with attention and processing, known as affective attention (Uusberg et al., 2013). Essentially, as Kissler, Herbet, Winkler and Junghofer (2009) note, "stimuli that people regard as emotionally arousing obtain prioritized processing" (p. 75). This phenomenon is believed to be a survival mechanism to ensure appropriate response to both negative and positive stimuli. Within this theory, personalized information about one's learning would be likely to receive more attention than general information, which was observed in this study. Combining cognitive load theory and affective attention theory suggests why suggestions and planning were self-reported for more time/attention than the figures. To an extent individuals are conscious of the time and effort they devote to something because they can sense greater and lesser cognitive load.

Finally, building on affective attention theory, as external judgments of one's competency hit hard at one's personal ego and self-efficacy (Ertac, 2011), it can be argued that figures depicting achievement are likely to be particularly arousing and therefore impactful –observed in this study

through the prevalent long-term recall of figures. The findings also reflect Smither, Bett, and Atwater's (2008) work on employees' recall of feedback wherein employees were likely to recall supervisor or external feedback – once again, external, authoritative judgements made a substantial long term impact on individuals. That the personalized suggestions for learning and planning opportunity were secondarily also recalled, and non-personalized information was not recalled at all, indicates the depth of impact that personalized feedback can have, as noted above.

5.2 Discussion 2: Processing of feedback report content

5.2.1 Cognitive and affective processes when receiving feedback on second language learning

Very little research exists on the processes by which language learners process feedback on their language learning, either cognitively or affectively. In the results reported for Theme 3, this study noted that several themes arose through qualitative self-report. Two affective themes were dealing with negative results and emotionality of response to the report content, and three cognitive themes were extent of relating report content to one's own life, dealing with comprehension challenges, and extent of critical evaluation of report content. The relevance of these themes, and purpose of their identification, is that they are all self-regulated learning strategies, but each one is particularly salient to processing of feedback.

Sample size prevents drawing many conclusions about how each of the themes may or may not typically operate alongside each other, but some inferences can be drawn from extant research on related topics. In particular, the types of metacognitive strategies used by participants while processing the report are informative for understanding which strategies are most relevant to feedback processing.

In terms of cognitive and metacognitive self-regulation strategies, the three strategies reported by participants in this study are not typically directly mentioned in taxonomies of second language learning strategies (Nosratinia, Ghavidel, & Zaker, 2015). However, similar strategies have been shown to be positively associated with independent and successful learning. For example, a similar strategy to relating report content to one's own life is inferencing, and inferencing skills are believed to be higher order thinking skills associated with successful reading in first and second languages (Grabe, 2009; Lee, 2013), with critical thinking (Magno, 2010), and with successful embedding of new information into long term memory and learning (Lau & Chan, 2007). Likewise, dealing with comprehension challenges proactively is a recognized reading

comprehension skill in first and second language (Comer, 2012), and problem solving skills in general are considered a key 'twenty-first century skill' (Scott, 2015). Finally, critical evaluation of content can be compared to critical thinking skills (Scott, 2015), again recognized as a strong metacognitive skill both in terms of evaluating content (Ku & Ho, 2010) and evaluating one's own progress (Pintrich, 2004), both of which related to critical evaluation of feedback report content.

However, the clearest finding regarding cognitive and metacognitive strategies is the confirmation of existing work on the relationship between language proficiency and cognitive and metacognitive strategy use. The evidence in other work clearly highlights that individuals with higher skill proficiency are able to utilize more cognitive and metacognitive strategies, and do so more effectively, than lower skill proficiency individuals (Lau & Chan, 2007; Lin & Yu, 2015; Zhang, Goh, & Kunnan, 2014). In this study, all four higher proficiency participants either did not report comprehension issues or worked to self-resolve them, whereas six of the seven lower proficiency participants exhibited helplessness in the face of comprehension challenges. Similarly, three of the four higher proficiency participants engaged in critical evaluation of report content, approving of some aspects and rejecting other parts. In contrast, four of the seven lower proficiency participants reported no critical evaluation, while two only reported limited evaluation cumulating in agreement. Clearly, the higher language proficiency participants were more able to utilize effective cognitive reading strategies and metacognitive self-regulation strategies to maximize the benefits of the report.

That said, it is notable that the third metacognitive strategy, relating the report content to one's own life, did not appear to be related to language proficiency. A possible explanation for this is that dealing with comprehension issues requires a certain level of language proficiency (Haastrup, 1991), while critical engagement with report content requires the confidence to do so (Anam & Stracke, 2016), as well as competency to deal with the demands of cognitive load (Kalyuga, 2009). Lower proficiency language learners may feel unqualified to critically engage with the report content, given that it comes from a place of linguistic authority, as well as struggle to free up enough cognitive space to do so.

In contrast, relating the report to one's own life can be done without challenging report content in any way and with limited aspects of the report. Moreover relating information to the personal

self rather than external abstract concepts has been shown to be cognitively less demanding (Snow, Burns, & Griffin, 1998). Therefore, it is hypothesized that relating the report to one's own life is cognitively less challenging than critical evaluation of content because it is relating information to one's personal experience rather than to abstract concepts of proficiency and ideal feedback content.

In terms of affective self-regulation strategies, among the sample of 11 interview participants, two neither minimized nor reacted emotionally to their results. Four participants minimized the significance of the results, three reacted emotionally, and two both minimized the results' significance and reacted emotionally. Minimizing or rejecting negative results may be an affective strategy to preserve a positive sense of self (Reb & Connolly, 2009), while emotional regulation is known to be associated with maintaining motivation in language learning endeavours requiring sustained effort or experiences of failure (Falout, Elwood, & Hood, 2009; Rose & Harbon, 2013). It would therefore seem that both affective strategies are maladaptive in terms of promoting long-term success in skill development.

Evidence from this study for this hypothesis is that four of the six participants who minimized their negative results reacted helplessly to comprehension problems compared to two of the five participants who did not minimize their results. However, two of the five participants who reacted emotionally both reacted helplessly in the face of comprehension problems while reading the report and engaged in no critical evaluation of the report content, while the other three engaged in critical evaluation that drew both positive and negative conclusions about report content, and either worked independently to resolve comprehension issues or did not recognize that they had comprehension issues. Moreover, these latter three participants all had high English language proficiency, whereas the former two had relatively low English language proficiency.

These findings indicate that minimizing results and reacting emotionally may be maladaptive self-regulation strategies for processing feedback, but that neither strategy is a fundamental barrier to achieving second language proficiency. Indeed, minimizing negative results may be a means of maintaining motivation (Pilipovic & Glusac, 2016, Reb & Connolly, 2009), and therefore facilitating future success, by ensuring that one's self-efficacy does not take too much of a direct hit, as typically happens when learners receive negative feedback on their skills (Ertac,

2011). However, Reb and Connolly (2009) note that consistent rejection of negative feedback usually becomes a barrier to learning.

Likewise, regulation of attention has been shown to be related to emotional regulation (Wadlinger & Isaacowitz, 2011), with the avoidance of distressing stimuli being associated with greater emotional regulation. Therefore in this feedback context, in which learners are supposed to attend to potentially distressing stimuli (poor results), emotional reactions may be evidence of attention to difficult-to-accept information. It is perhaps therefore the subsequent negotiation of these difficult emotions and this challenging information that determines the effectiveness with which the learner engages in feedback.

5.2.2 Affective and cognitive outcomes after processing feedback on second language learning

Participants' responses to processing outcomes were generally positive in terms of trust in report content, reflection on one's English skills, and desire to use the report. This coheres with existing research that indicates learners are predisposed to liking personalized feedback about their learning (Orsmond, Merry, & Reiling, 2005; Walker, 2009). The finding also indicates that the learners in this study might be more likely to use the feedback (Carless, 2006; Corbalan, Paas, & Cuyper, 2010; Lee, Lim, & Grabowski, 2010). Unfortunately, many participants also felt in need of help to use the report, likely substantially decreasing the chances that they would use their reports in learning without support. While this phenomenon is not uncommon (Cramp, 2011; Rae & Cochrane, 2008), the sense of needing help, as will be seen later in the discussion, most certainly impacted how people planned to use the report.

Regarding processing a feedback report, perhaps the two most crucial outcomes are a desire to use it and not feeling in need of help to do so. Indeed, the latent class analysis results indicated that in terms of differentiating learner processing outcome profiles, the participants in this study differed mainly according to desire to use the report, and feeling in need of help to use it. On average, a strong desire to use the report was associated with high levels of reflection on one's English skills, and relatively high levels of trust in the report content. In contrast, strongly feeling in need of help to use the report was on average associated with relatively high feelings of overwhelmedness. These findings are intuitively appealing, but also concur with earlier work that shows learners who view feedback as useful, comprehensible and trustworthy are more likely to use – or in this case want to use – that feedback (Furnborough & Truman, 2009;

Pokorny & Pickford, 2010; Timmers & Veldkamp, 2011). However in this study, a nuance between desire to use the report and feeling in need of help to use the report emerged as an area of significant difference among learners. This difference will be shown to be significant in later discussion, in terms of how learners intended to use their feedback.

Moreover, the results of matching reported cognitive, metacognitive and affective strategies with processing outcomes, although based on small sample sizes, do appear to match existing work in cognitive processing and language learning. As such, the results begin to create a profile of the types of processes that language learners use while processing feedback on their language skills, particularly in terms of their desire to use the report, and their feelings about needing help to use it.

First, participants who reported no comprehension difficulties ($n = 2$) and reported no critical engagement ($n = 4$) had much lower mean desire to use the report than other participants.

Moreover, those reporting no comprehension difficulties also reported relatively low mean trust and reflection, demonstrating the relationship between desire to use the report, trust and reflection. Therefore, while no comprehension difficulties and acceptance of the report might appear positive outcomes, this study's results indicate that in fact the absence of comprehension difficulties and critical engagement may indicate lack of engagement with feedback. This finding complements Fernandez-Toro and Hurd's work (2014) modelling how language learners actively engage with feedback in order to use it, and Orsmond, Merry, and Reiling's (2005) work on how university students use feedback, showing that cognitive and emotional negotiation with feedback is in fact a prerequisite to engagement with the information that feedback contains.

Moreover, participants who sometimes related the report to their own lives ($n = 4$) – as opposed to frequently doing so or not doing so at all, and participants with only positive evaluations of report content ($n = 3$) – as opposed to those who reported no evaluation or reported critical evaluation, on average strongly felt in need of help to use the report. In addition, those who made no mention of needing help ($n = 2$) reported low mean feelings of needing help. The implication is that engagement is indeed necessary, but that the quality of processing and engagement may be indicative of learners' confidence that they can use their feedback. It should also be noted that participants reporting only positive conclusions about the report also had the highest mean trust in report content and reflection on their English skills. Jang, Dunlop, Park and van der Boom

(2015) made a similar finding, noting that young learners displaying positive psychological learning characteristics such as strong mastery goal orientations were also more likely to critically engage with the feedback than fully accept it.

Finally, the earlier discussion on the emotional benefits of apparently maladaptive approaches to dealing with undesired feedback is furthered by the findings on overwhelmedness. In this study, participants who did not minimize their personal responsibility for the scores ($n = 5$) reported greatest feelings of overwhelmedness, suggesting that deflecting responsibility for results may have a beneficial processing effect, although Reb and Connolly (2009) note that consistent deflection decreases the probability of changing behaviour appropriately. However in the context of this study, the finding is particularly interesting given that deflecting responsibility did not appear to be associated with differences in wanting to use the report, trust in the report content, or reflection on one's English skills, all desired processing outcomes. Potentially, deflecting personal responsibility might be conceived as a means of critical engagement with the report, if accompanied with positive engagement outcomes such as reflection on one's English skills.

In summary, a tentative model of processes by which highly motivated learners such as adult immigrants engage with feedback on their second language skills is proposed in Figure 23.

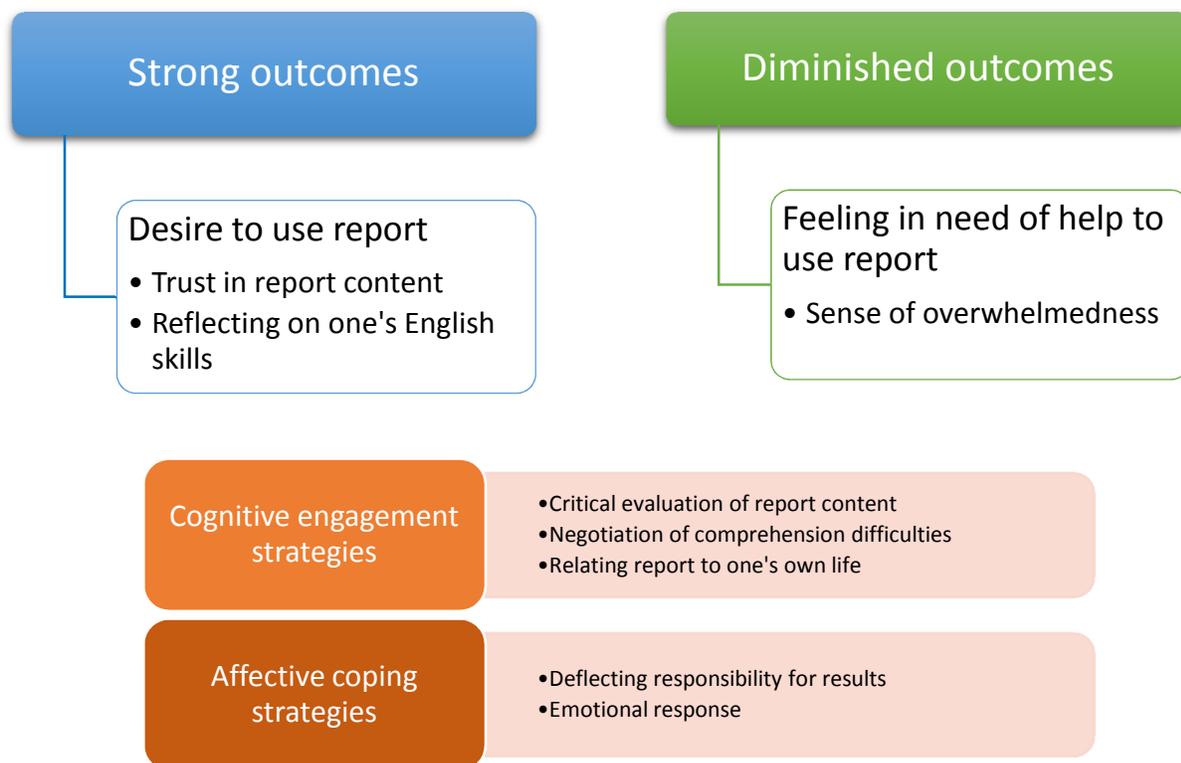


Figure 23. Key and secondary processing outcomes, with processing strategies that support stronger outcomes

5.3 Discussion 3: Factors that mediate usage of feedback on second language learning

There has been very little research conducted on what usage of feedback by language learners actually looks like, and the individual characteristics affecting said usage. Therefore, this section of the discussion will principally draw on existing literature of the psychological characteristics observed to associate with various forms of usage, to discuss ways in which they appear to be interacting with usage. Table 37 summarizes the results from Theme 4 regarding the significant statistical relationships observed between feedback report usage and individual characteristics, reported attention, and reported processing outcomes. In addition, the delayed recall interviews provided qualitative data, and the findings from Theme 1 provided much of the material to explain differences among learners.

Table 37

Significant relationships between initial report usage and background, attention and processing variables

	M	PP	PA	fix	totscor	SA	descrp	fig	sug	2much	trst	thnk	wantuse	help
moreskill							+		+					
moresugg					+									
#goal							+							
#plan			-		-									
planspec		-					+					+		
#mon					-						+			
onintmon				-										
oneexmon	+				+									
anymon	+			-						-				
teacher	+				-									
imprvEng		-			-									+
whatskill								+		-	+			+
reportsay				+		+								
levelEng						+								+
SAvPPM		+					+					+		
otherpthink		+									-			
hwdisgre	+												-	

Individual characteristics variables: M = mastery orientation; PP = performance prove orientation; PA = performance avoid orientation; fix = fixed beliefs about intelligence; totscore = total raw test score; SA = self-assessment composite score

Reported attention variables: descrp = skill descriptions; fig = figures; sug = suggestions for learning

Processing outcome variables: 2much = overwhelmedness; trst = trust in report content; thnk = reflecting on one's English skills; wantuse = desire to use report; help = feeling in need of help to use report

Usage variables (planning opportunity): moreskill = selecting more skills to work on; moresugg = selecting more suggested skills to work on; #goal = number of goals written; #plan = number of action plans written; planspec = specificity of action plans; #mon = number of monitoring plans written; onintmon = at least one internally-monitored monitoring plan; oneexmon = at least one externally-oriented monitoring plan; anymon = at least on substantive monitoring plan (either externally or internally oriented)

Usage variables (planned discussion partners/topics): teacher = intention of talking to English teacher; imprvEng = How to improve my English; whatskill = What areas of my English skills I should focus on; reportsay = What the report says about my English language skills; levelEng = My level of English proficiency; SAvPPM = How my self-assessment compares to the test results; otherpthink = What the other person thinks about my English skills; hwdisgre = How much I agree or disagree with the report

5.3.1 Motivation and environment

A key mediator of the feedback report presented in this study was the deep motivation of the participants to master English for their own life purposes, and the extent to which they utilized

their socially accessible resources, particularly spouses and children, to reach those goals and garner support along the way. The literature is very clear that motivation is a key engine in persistent learning (Dörnyei, 2005, 2009; Duguay, 2012; Gu & Cheung, 2016; Han, 2009), and that learners adapt the messages in feedback for their own purposes (Orsmond, Merry, & Reiling, 2005; Derwing & Waugh, 2012). The strategies implemented by this study's participants are evidence of how their motivation takes practical form in terms of language learning feedback usage.

In addition, a classroom effect may have impacted this study. Learners with fixed beliefs about intelligence but in a mastery-oriented environment are more likely to deal well with criticism and challenge (Thompson & Musket, 2005), and in this study there were two classrooms where participants' responses to feedback were remarkably consistent. In one classroom, participants were highly accepting of the report content despite disappointing and surprising results, and despite many participants holding fairly strong fixed beliefs about intelligence and performance prove orientations. In this program, the classroom teacher was strongly supportive of the study and the learning environment can be described as bright, cheery, communal and relaxed. In another program, the report was politely but collectively questioned and eventually rejected by both the teacher and the participants because it contradicted the in-class ongoing program of assessment. This class was also a strong, supportive community, but was much more focused on particular assessments and measures of observable progress.

It could be argued that the second program had too high a focus on assessment, and as Shute (2008) argues, creating an environment too focused on assessment is destructive for learning outcomes. However, there is also ample evidence of the benefits of goal setting and monitoring learning (Hattie & Timperley, 2007), and these principles were likely more comprehensively implemented in the second program (although classroom environments were not an area of study in this research). Therefore, rather than reflecting the specific characteristics of a classroom, the group reactions observed may simply reflect the very important role that the learning environment has on processing and usage of feedback (Poulos & Mahony, 2008; Rae & Cochrane, 2008).

The role of context as a key factor in successful feedback usage was also indicated in the delayed recall interviews, where participants co-opted their social resources, both in terms of language

practice through television, children's homework, and community events, and in terms of support, particularly using family members to mediate challenging language environments. Conversely, those participants whose home circumstances were sufficiently inhospitable to English study, for example heavy childcare and household burdens, travel to country of origin, or employment and finance stresses, were unable to utilize feedback in their learning – or indeed learn much in general. Derwing and Waugh (2012) discuss the power of creating conditions where immigrants can make space to learn language and cultural skills necessary for integration and adaptation in their new home, and this study also makes this finding.

5.3.2 Goal orientation and beliefs about intelligence

Goal orientation and beliefs about intelligence emerged in the study as regularly associated with usage of the report. There is a theoretical justification for this, as goal orientation has been shown to affect feedback seeking behaviour (Jang, Dunlop, Park, & van der Boom, 2015; Pappachan, 2008; VandeWalle, 2003), and using the report was a form of feedback seeking. This study did not add directly to evidence that language learners with performance orientations seek less feedback, such as found by Butler (1993) and Mantou Lou and Noels (2016), because the study gave no opportunity for participants to avoid feedback.

However, participants with higher performance prove orientations were more likely to want to discuss with others about the other person's opinion of their English skills, and to discuss how the self-assessment compared to the test results. This finding is similar to that of Jang, Dunlop, Park and van der Boom (2015), in which performance oriented learners focused on performance outcomes, although it is unclear whether participants wanting to discuss these topics were accepting the results or not. However to inform this gap, this study found that learners with stronger mastery orientations were more likely to want to discuss how much they agree or disagree, reflecting the finding by Dunlop, Park and van der Boom in which mastery oriented learners were more likely to critically engage with feedback.

Note also that a performance prove orientation was not an entirely unhelpful attribute; participants holding strong performance prove orientations were more likely to want to discuss how to improve their English, particularly if they had low test scores. However, they were also likely to write less specific action plans, and performance avoid learners were likely to write fewer plans at all. It is perhaps significant that although performance oriented learners are

focused on their results, they may be less focused on making the plans that will help them improve. Such a finding may be related to Tuckey, Brewer and Williamson's (2002) finding that learners with performance orientations are less interested in information that is useful for their learning.

In contrast, Gorges, Kandler and Bohner (2012) found that a mastery goal orientation facilitated positive attitudes to language learning regardless of ability, and Vandewalle (2003) observed that mastery oriented learners were more likely to seek feedback. This study built further on these observations, finding that mastery oriented learners were more likely to want to talk to their teacher about the report, and they were also more likely to write an externally-oriented monitoring plan, in which they sought external feedback on their language learning. It can therefore be proposed that learners with stronger mastery oriented learners are indeed more open to external feedback.

In terms of beliefs about intelligence, learners with stronger fixed beliefs were less likely to write an internally regulated monitoring plan, or indeed any real monitoring plan, and more likely to want to discuss what the report said about their English skills. Indeed, Mantou Lou and Noels (2016) also found that language learners with fixed beliefs about intelligence were likely to respond helplessly to feedback, and failure to make substantive monitoring plans may be evidence of this tendency. Focusing on what the report says about their English skills may likewise reflect these participants' beliefs that test results are a reflection of aptitude rather than current ability (Dweck, Chiu, & Hong, 1995).

A final point of discussion concerns the existence of multiple goal orientations at one time (Button, Mathieu, & Zajac, 1996; Jang, Dunlop, Park, & van der Boom, 2015). As shown in the latent class analysis of typical goal orientation profiles, participants in this study held multiple goal orientations to varying degrees. However, in the analyses for feedback usage, only one goal orientation at once was ever significantly associated with a feedback usage behaviour, and always in conjunction with other characteristics such as language proficiency, attention and processing outcomes. This finding indicates that while language learners may latently hold multiple goal orientations during language learning, these orientations emerge as significant in learners' usage of feedback at particular times, for example when self-concept is contradicted by a valued authority source such as a test, when they distrust the report content, or when they are

feeling confident about their ability to use the report. This finding thus adds to the evidence that goal orientations are not direct predictors of usage, just as they are not direct predictors of academic achievement (Elliot & Church, 1997; Mantou Lou & Noels, 2016), but rather latent aspects of learners' psychology that impact learning behaviours such as feedback usage.

5.3.3 Language proficiency and self-assessment

English reading proficiency was earlier shown to have a strong effect on how participants processed information, and there were also some findings regarding the impact of participants' English reading proficiency on report usage. Firstly, regardless of other individual characteristics, participants with lower test scores were more likely to want to discuss the report with their teacher, and to discuss how to improve their English. This finding is intuitively logical and reflects the value that learners place on discussing their skills with their teachers and other trusted people (Carless, 2006; Rae & Cochrane, 2008).

Secondly, higher English proficiency participants were much more selective in their use of the planning section of the report. Higher English proficiency participants were less likely to write several action or monitoring plans but more likely to select an external monitoring plan, and they were also more likely to use a skill suggested for practice on the report, in their plans. Although there is no research on how higher and lower language proficiency learners use feedback differently, other research has found that more proficient learners are able to apply cognitive strategies in greater variety and with more selectiveness (Saengpakdeejit & Intaraprasert, 2014; Lin & Yu, 2015). The differences observed in this study can be interpreted as manifestations of this greater cognitive dexterity, with higher proficiency learners demonstrating more selectivity in their plan writing.

Finally, the fact that higher proficiency participants were more likely to select suggested skills is more a reflection of choices made by lower proficiency participants rather than those with higher proficiency. Most lower test score participants selected "Using vocabulary" as a skill, and ignored the recommendations provided to lower proficiency learners that focused on the easier skills (in this test) of pragmatics ("Using culture") and main ideas ("Distinguishing information"). This finding is an example of agency among participants, who engage with report content selectively based on their own interpretations of their issues (Orsmond, Merry, & Reiling, 2005).

In terms of the role of self-assessment in feedback, discrepancy differences were observed in this study between lower- and higher- English proficiency participants that indicated the widely recognized Dunning-Kruger effect (Schlosser, Dunning, Johnson, & Kruger, 2013), wherein lower proficiency participants grossly overestimate their proficiency, and higher proficiency participants are relatively accurate estimators of their proficiency. This is a natural psychological phenomenon in all areas of expertise and learning, but one that is important to account for and monitor when providing feedback to learners.

Previously, Jang, Dunlop, Park and van der Boom (2015) noted that young learners' attention was focused by including self-assessment alongside test results, a position supported by Alderson, Brunfaut and Harding (2014). In this study, while discrepancies between total score and self-assessment were not associated with any types of report usage, higher self-assessment score was associated with desire to discuss two conversation topics: what the report says about one's English skills, and one's level of English. Given that the majority of participants overestimated their English reading proficiency, both of these topics are conceivably related to shock regarding large differences between test results and self-assessment.

Therefore in this study the self-assessment does seem to be serving as an indicator to focus language learners' attention on their ideas about English reading proficiency. This finding is confirmed in the interviews, in which participants regularly commented on discrepancies between test results and self-assessment, and how they seemed better or worse at English reading than they had previously realized. There is however a note of caution in this outcome; Ertac (2011) found that larger-than-warranted drops in self-image could be caused by negative feedback among all but the most confident learners, and Hattie and Timperley (2007) were wary of negative effects such as decreased motivation. This phenomenon was not directly assessed in this study, and participants' motivation to master English generally appeared to carry them through the disappointment many of them felt. However, at least one interview participant – with relatively low English language proficiency, relatively high performance avoid goal orientation, and low self-efficacy – noted that she kept the report at home due to shame, and did not think about it. As such, the grounding that self-assessment provides may be too bruising for some learners with at-risk psychological profiles.

5.3.4 Attention, cognitive processing and affective processing outcomes

The attentional and cognitive processing that takes place when highly motivated adult second language learners such as those in this study receive feedback on their language skills has already been discussed. The ways in which these processes interact with feedback usage is now discussed. Very little empirical work has been done in this area, so the results are discussed here in the context of the theoretical works of Pintrich (2004) and Hattie and Timperley (2007).

First, Pintrich's (2004) framework for self-regulated learning identified that time and effort are a key area of regulation for Pintrich's first phase of self-regulation, 'forethought, planning and activation'. It can therefore be argued that evidence of effort in planning, such as greater specificity or more substantial plans, indicates positive self-regulation. Moreover, attention is related to depth of cognition (Duchowski, 2007; Wadlinger & Isaacowitz, 2011). Therefore in this study, a relationship should be observable between attention to the report, and quality of planning.

Reported attention to the introduction and the planning section were not significantly associated with any feedback usage behaviour. While the lack of relationship between introduction and usage is somewhat intuitive given its non-personalized format and the known impact of personalized feedback (Hattie & Timperley, 2007; Orsmond, Merry, & Reiling, 2005; Walker, 2009), it is interesting that reported attention to the planning section was not related with any planning behaviours.

In fact, the report section most often associated with feedback usage, including planning behaviours, was attention to the skill descriptions. This was the section that relatively fewer participants reported spending much time on. Participants who reported more attention to skill descriptions chose more skills to work on and wrote more goals, and wrote actions plans with greater specificity. They were also more likely to want to discuss how the self-assessment compared to the test results. It is unclear why attention to skill descriptions should be particularly relevant, but there may be an influence of language proficiency, as attention to skill descriptions was associated with higher language proficiency due to the dense linguistic nature of the information provided in this section. If this is the case, attention to skill descriptions can be considered to reflect both cognitive effort and the ability to enact this effort (due to sufficient language proficiency). The link between usage and attention then becomes relevant, and fits with

Pintrich's (2004) framework for self-regulated learning. Thus we see that more goals and greater plan specificity are related to effort in attending to the report, contingent on comprehension.

This proposal is supported by the finding that more reported attention to learning suggestions was also associated with selecting more skills to work on, with an interaction effect between skill descriptions and suggestions. Suggestions received much attention even among lower English reading proficiency participants, and may therefore be operating as the clearest indicator of effort investment from lower English proficiency participants. Finally, more reported attention to figures was associated with greater desire to discuss what skills to focus on, and adds to the argument that learners use figures as a framework to structure their reflection.

However, these were the only instances of figures and suggestions being related to feedback usage. This relatively sparse set of relationships may be due to weaknesses in the measures of attention and usage, and/or may reflect Hattie and Timperley's claim that multiple factors are in play during reception and processing of feedback, and that other factors were simply more important. In this case, Hattie and Timperley's theories are somewhat confirmed by the results of this study, in that goal orientation, beliefs about intelligence, and English language proficiency were more consistent predictors of usage than attention.

Just as attention had a relatively weak relationship with usage of the planning section compared to individual characteristics, processing outcomes also had a relatively weak relationship. Plan specificity was associated with more reflection on one's English skills, once again reflecting the relationship between effort investment and planning. Higher levels of trust in the report content were associated with more monitoring plans, while greater sense of overwhelmedness was associated with fewer substantive monitoring plans.

However, processing outcomes were more consistently related with intended discussion topics. This finding indicates that processing outcomes do have an impact on how language learners intend to use the feedback they received, and supports Hattie and Timperley's argument that feedback must support positive processing outcomes. In particular, feeling in need of help to use the report was associated with three out of seven discussion topics. Those feeling in greater need of help were more likely to want to discuss their English level, what skills to focus on, and how to improve their English. In contrast, desire to use the report was related to only one topic; participants with less desire to use the report were more likely to want to disagree how much

they agreed or disagreed with the report. Neither of the outcomes were related to usage of the planning opportunity. However note that desire to use the report was highly negatively skewed in this study, reflecting the motivation of the participants on their wider language learning journey, but leaving less information to draw on for the implications of not wanting to use the report.

Referring once again to Hattie and Timperley's model of feedback, providing information that the learners want and need to move forward is of paramount importance in feedback delivery. In this respect, the conversation topics associated with feeling in need of help are perhaps best interpreted as forms of help seeking, and provide information on the type of information sought by learners. VandeWalle (2003) and Jang, Dunlop, Park and van der Boom (2015) note that goal orientation has an impact on help seeking behaviour, but goal orientation was only significantly associated with one of these three topics. Instead, more complex profiles of learners seeking specific advice emerged. For example, discussing what skill to work on appeared to be a concern of participants who were positive about the report and not overwhelmed but wanted more information. In contrast, those with high self-assessments and feeling in need of help wanted to discuss their level of English.

Additionally, trust in feedback content was named by both Hattie and Timperley (2007) and Nicol and Macfarlane-Dick (2006) as a primary base on which other processing is built, and Furnborough and Truman (2009) also showed this empirically for language learners. In this study, greater trust in report content was associated with higher probability of intending to discuss what skills to focus on and lower probability of discussing what the other person thought of one's English skills. This finding indicates that in practice, trust in the report content meant an increased probability of desiring to use external sources to build on the information drawn from the feedback report rather than to seek a second opinion on the original information.

Finally, greater reflection on one's English skills was related to increased probability of wanting to discuss how the test results compared to the self-assessment. Reflection is a foundational activity in self-regulated learning activities (Hattie & Timperley, 2007; Pintrich, 2004). It is notable that both instances in which more reflection on one's English skills was significantly related to an aspect of increased report usage, more attention to skill descriptions was also significantly related. In addition, performance prove orientation was significant, although the direction of the relationship varied. This finding is further evidence that reflection is a form of

cognitive investment, and that goal orientation mediates the amount of investment that languages learners are willing to make in language learning opportunities (Gorges, Kandler & Bohner, 2012; Nakayama, Heffernan, Matsumoto, & Hiromori, 2012; VandeWalle, 2003).

Chapter 6: Conclusions

6.1 Summary of results

Most existing research on educational feedback considers feedback in the context of classroom instruction or computer-based learning environments. In contrast, this study addressed how English language learners use computer-based feedback on their English language proficiency delivered outside a classroom or computer-based learning context, although a majority of participants were also studying English in a classroom. Therefore, the findings and implications of this study centre around the demands of providing automated proficiency feedback outside the classroom, without a teacher mediator.

Based on the findings in this study, the road to using language learning feedback is clearly a complex one. Many researchers have previously outlined the various relationships that might be interacting, including Hattie and Timperley's (2007) general model of feedback, Clark's (2012) argument that assessment and self-regulated learning as inextricable, Wagner's (2015) framework of classroom-based second language writing feedback, and Fernandez-Toro and Hurd's (2014) model of factors for feedback engagement for language learning. However the focus has generally been on factors affecting feedback usage, and outcomes of feedback, rather than the processes that language learners enact while receiving feedback. In contrast, this study attended mostly to the processes language learners used to deal with feedback, then looked at how usage related to attention, processing and individual characteristics.

This study found that attention to report sections varied. While the most time was spent on the text-dense sections, the most attention was given to personalized sections, particularly those that carried information about how to move forward and improve. In terms of what stayed with study participants longest, the figures had strongest impact, although details were not necessarily recalled. Self-assessment, as predicted by Alderson, Brunfaut and Harding (2014) and previously noted by Jang, Dunlop, Park and van der Boom (2015) was one of the reasons the figures had such strong impact. Implications are that multiple aspects of feedback as useful at different stages in feedback processing and usage.

The study also found some methodological implications for eye tracking. There was an almost exactly inverse relationship between which parts of the report were recalled one month later, and

which parts of the report had highest fixation times during reading. There was also a mostly reverse relationship between what participants reported spending time looking at compared to what the fixations said they looked at. The discrepancies do not appear to originate from reporting inaccuracies by either participants or the eye tracker, but that each measure was tapping into different forms of attention, with subsequent implications for the different types of associated cognition that took place. This finding is relevant for researchers exploring attention, and each method is useful for distinct purposes. An additional finding was that Owen's (2016) argument that eye tracking can prompt limited recall when used on its own is justified, so eye tracking is best used as a prompt to further explore cognition after recall takes place.

Another main finding of the study was the specific cognitive and affective processes used by the language learners when dealing with feedback. Processes and strategies included emotional responses, deflecting responsibility for results away from themselves, critical evaluation of report content, negotiation of comprehension difficulties, and relating the report to one's own life. The two affective strategies observed are typically considered maladaptive learning strategies, as refusal to accept difficult information reduces the ability of an individual to change their behavior (Reb & Connolly, 2009), and emotional responses increase the chances of reacting inappropriately to information (Wadlinger & Isaacowitz, 2011). However, in this study the strategies appeared to have mixed roles, and it would seem that the combination of strategies is more important than the individual list of strategies used.

Regarding use of the feedback report, a key factor for participants in this study was simply the opportunity and energy to do so. The importance of the social context in which learning takes place is known to be very key for many aspects of learning (Dörnyei, 2005; Vygotsky, 1978; Hattie & Timperley, 2007), and this study's findings regarding the limited usage of the reports one month later highlight the limited impact that feedback outside of learning programs can have.

That said, the feedback reports were generally well received in that a clear majority of participants wanted to use them, trusted the content, and intended to discuss various issues related to their English studies as a result of receiving the report. Moreover, most delayed recall interview participants who wanted to use their report and had the opportunity to do so, did so. However a relevant finding is that the participants all co-opted the report suggestions and planning according to their own situations. So those who regularly watched television with their

children, decided to do so with more focus. Those who regularly read free newspaper decided to pay attention to vocabulary more systematically.

These findings highlight that suggestions and resources included in feedback need to focus on the resources available to language learners, otherwise they will not be used. Moreover, advice on next steps is best designed to include easily accessible and preferred activities, and perhaps offer guidance on extending the effectiveness of these activities for language learning purposes. While this is not new knowledge – any skilled classroom teacher would confirm the veracity of these statements, it is a finding of this study, and worth repeating.

Finally, this study found relationships between feedback usage and attention, processing outcomes, and individual characteristics such as goal orientation and English language proficiency. Based on these findings, a set of key factors mediating how language learners use non-teacher-mediated feedback is proposed in Figure 24.

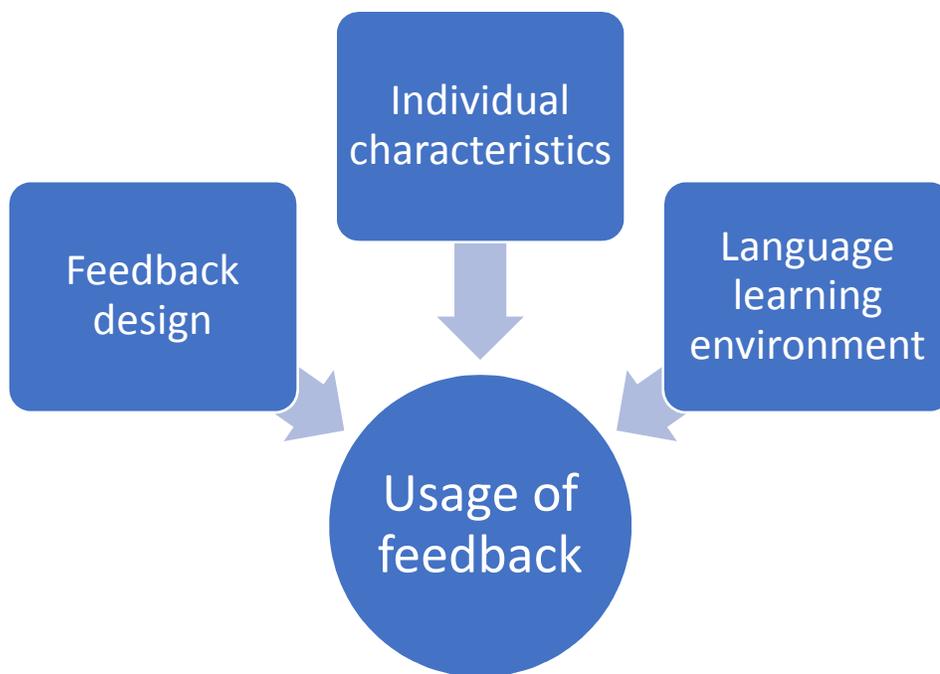


Figure 24. Key factors mediating how language learners use non-teacher-mediated feedback. As already documented (cf. Clark, 2012; Fernandez-Toro & Hurd, 2014; Hattie & Timperley, 2007; Wagner, 2015), the design of feedback strongly impacts usage of feedback. Design choices might include language of feedback, amount of text used and number of skills reported, use of features such as test scores, figures, self-assessments, peer assessments, comparisons with other

learners, extent of personalization, and use of scaffolding for learners with weaker self-regulated learning skills. Additionally, the individual characteristics of the language learner have a large impact on how feedback is used; language proficiency is particularly relevant given that much information between people is transmitted via language of some kind. However, motivation is another strong mediator, as well as underlying characteristics such as self-efficacy, self-regulated learning skills, and orientations toward learning and intelligence. A final key factor on usage of feedback is the language learning environment, for example a classroom, a computer-based learning environment, or outside formal programming; within a community that speaks the target language – or not. According to how these aspects are set up, language learners will interact with feedback in different ways, and eventually use it (or not), with greater or lesser impact on learning.

6.2 Limitations of the study

One of the principal limitations of this study lies in the distinctive characteristics of the population participating in the study. Adult immigrant second language learners, despite the many daily challenges they face, are in an enviable position in terms of maintaining motivation for second language learning. Their relatively immersive environment also provides much support for language practice. However, many second and foreign language learners are not in such conditions, and thus the distinctiveness of this population makes generalization to populations such as learners in foreign language classrooms tenuous. Additionally, due to proactive government policy and societal openness to newcomers, Canada's immigrants may currently be relatively unghettoized and unmarginalized in comparison to many immigrant communities worldwide. Therefore, the findings of this study may not even be applicable to immigrant communities in other countries where exclusion and lack of support for gaining social participation are more entrenched.

That said, many of the principles about the ecological characteristics of feedback observed in this study, such as the impact of goal orientation, motivation, language proficiency, and environment have been observed in other populations. However, the relative impact of factors will likely vary according to the context of a population. Moreover, the skewed nature of several observed variables, such as desire to use the report, reduced the ability of the study to identify profiles for less engaged and motivated learners.

Another important limitation of this study was that feedback was delivered in English, and participants also wrote their plans in English, although they were not overtly instructed to do so. Moreover during interviews, most participants had to use English, although Mandarin Chinese speakers were able to use Mandarin when they wished. Given the moderate English language proficiency of many participants, using English most certainly impacted not only on their processing of report content, but also limited participants' ability to articulate plans and thoughts during planning and interviews, and understand the surveys.

Another limitation of the study was the measures used, which due to logistical constraints were unable to be piloted and validated prior to data collection. Therefore, some poorly performing items were dropped and not replaced, likely reducing the sensitivity of the instruments in detecting constructs such as processing outcomes and beliefs about intelligence. In addition, the decision was made to collect cognitive and affect processing through qualitative methods, in order to avoid predetermining the processes. However, a reduced final interview sample size made drawing wider conclusions another tenuous endeavour.

Finally, a limitation for the interviews was a methodological limitation for eye tracking. The eye tracker was unable to scroll and unable to admit data input, so the report was split into six pages, and the two pages containing the planning section had no writing facility. Therefore the eye-tracking and associated interview data were limited in that they could not exactly reflect the processes of participants receiving their report via the internet, which had scrolling on a single page, and the facility to type in plans.

6.3 Implications and suggestions

Several implications and suggestions can be derived from the findings of this study. First, there are implications for test design. If test design is to be oriented to effects (Fulcher & Davidson, 2003), outcomes (Bachman, 2005) and impact (Weir, 2005), and continued learning is a positive effect/outcome/impact of a test, it can be argued that providing feedback for learning purposes will improve test design even for high stakes tests primarily serving as verifications of second language ability. Such an approach might be called a feedback-based approach to test design, and would constitute a substantive link between second language test design and validation approaches, and assessment-for-learning approaches.

However, this study found that the path to using feedback for learning is complex, and not only depends on feedback design features, but also on the language learning environment as well as individual characteristics of the language learners. Therefore, it may be suggested that test design fully take into account the information that will need to be provided to language learners in order for them to use the test results for learning purposes. The features of this information will depend not only on the construct being tested and the testing purpose, but also the language learning environment and individual characteristics of the test takers. Relevant questions might be: What level of detail about learners' skills is required? What kind of information do learners expect and are already able to comprehend? What information can be provided that is new to learners, but they will be able to grasp? How can test design incorporate elements of self-assessment and/or peer assessment? Will scores be useful for learning, or does test output need to be interpretable into words?

Secondly, there are implications for those working with language learners, including those researching effectiveness of computer-based learning. Given the different ways that learners process and use feedback, it would seem there is a need to structure feedback, particularly where feedback that is provided without an expert mediator such as a teacher, so that learners with different characteristics, most notably second language proficiencies, self-regulated learning skills, and motivations for learning, are more able to use that feedback in their learning. Such an approach could free up classroom time to focus on language learning, which as a multi-faceted process of socialization and personal development that contains many other tasks than processing feedback. For those forced to study independently, such support might greatly improve learning outcomes, or at least learning persistence. While these implications certainly apply to traditional classrooms, the implementation of new technologies in this area might allow for wider application due to the labour-saving benefits of these technologies. However, it should be underlined that language learning is a fundamentally social process concerning learners' social identities, and any assumption that sufficiently personalized feedback would result in the removal of other people from the learning process, particularly regular, physical learning communities such as classrooms and domain experts such as teachers, misunderstands the dynamics of human language learning.

Finally, there are methodological implications for researchers intending to access language learners' perceptions of attention and processing, as this study found that at least with attention, different research methods tap into different aspects of attention and processing. Many empirical measures, such as eye tracking, measures of biodata and even facial recognition provide objective data, but processes that occur in the mind are often beyond the reach of these measures, and thus subjective measures designed to access unevaluated cognition such as think aloud and recall interviews, as well as surveys and interviews designed to collect self-report evaluations, are also valuable sources of data. Therefore a mixed methods approach is advocated for researchers aiming to gain a deeper understanding of processing including for feedback on language learning.

6.4 Areas for further research

This study observed how adult language learners with different characteristics interacted differently with computer-based feedback that was not teacher-mediated, and observed the impact this may have had on feedback usage. The most immediate further work required is to develop models for how those learners susceptible to less effective use of feedback can be supported to use feedback more effectively. An experimental study would be a valuable contribution to this topic. In addition, given the impact that figures had on participants in this study, a specific useful research topic would be the benefits and risks of using figures that describe performance, and how alternative methods of referencing performance might also benefit learners.

This study was focused on adult immigrant English language learners in Canada, and provided feedback in English due to linguistically diverse participant population and the resource limitations of the study. Therefore further research would also be helpful in investigating feedback processing strategies among language learners in other common language learning contexts, such as foreign language classrooms or those in classroom immersion programs where the community speaks other languages. In addition, research that demonstrates the differences in feedback processing and usage that take place when feedback is provided in a mother tongue or highly proficient language, compared to the target learning language, would also further the field substantially.

Finally, feedback in this study was delivered via computer technology and without a teacher mediator. It would be useful to conduct research that considers how the findings of this study can change when technology and a teacher are involved, and when only a teacher is involved.

References

- Alderson, J.C. (2010). "Cognitive diagnosis and Q-matrices in language assessment": A commentary. *Language Assessment Quarterly*, 7, 96-103.
- Alderson, J. C., Brunfaut, T., & Harding, L. (2014). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics*, 36(2), 236–260. <https://doi.org/10.1093/applin/amt046>
- Amaral, L., Meurers, D., & Ziai, R. (2011). Analyzing learner language: towards a flexible natural language processing architecture for intelligent language tutors. *Computer Assisted Language Learning*, 24(1), 1–16. <https://doi.org/10.1080/09588221.2010.520674>
- Anam, S., & Stracke, E. (2016). Language learning strategies of Indonesian primary school students: In relation to self-efficacy beliefs. *System*, 60, 1–10. <https://doi.org/10.1016/j.system.2016.05.001>
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38(2), 113-125.
- Awad, G. (2014). Motivation, persistence, and cross-cultural awareness: A study of college students learning foreign languages. *Academy of Educational Leadership Journal*, 18(4), 97-116.
- Aydeniz, M., & Southerland, S. A. (2012). A national survey of middle and high school science teachers' responses to standardized testing: Is science being devalued in schools? *Journal of Science Teacher Education*, 23(3), 233–257. <https://doi.org/10.1007/s10972-012-9266-3>
- Ayres, P., & Paas, F. (2012). Cognitive load theory: New directions and challenges. *Applied Cognitive Psychology*, 26(6), 827–832. <https://doi.org/10.1002/acp.2882>
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1–34. https://doi.org/10.1207/s15434311laq0201_1
- Baleghizadeh, S., & Masoun, A. (2013). The effect of self-assessment on EFL learners' self-efficacy. *TESL Canada Journal*, 31(1), 42-57.

- Ballard, L., & Lee, S. (2015). How young children respond to computerized reading and speaking test tasks. Conference presentation at *Language Testing Research Colloquium 2015*. Toronto: LTRC.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology; Washington*, 25(5), 729-735.
- Bangert-Drowns, R., Kulik, J. A., Kulik, C. C., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213-238.
- Bao, M., Egi, T., & Han, Y. (2011). Classroom study on noticing and recast features: Capturing learner noticing with uptake and stimulated recall. *System*, 39(2), 215–228.
<https://doi.org/10.1016/j.system.2011.05.001>
- Bartlett, L., Dowd, A. J., & Jonason, C. (2015). Problematizing early grade reading: Should the post-2015 agenda treasure what is measured? *International Journal of Educational Development*, 40, 308–314. <https://doi.org/10.1016/j.ijedudev.2014.10.002>
- Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, 30(4), 441–465. <https://doi.org/10.1177/0265532212473244>
- Biber, D., Nekrasova, T., & Horn, B. (2011). *The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis*. Princeton: ETS. Retrieved from <http://www.ets.org/Media/Research/pdf/RR-11-05.pdf>
- Bisson, M.J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye-tracking study. *Applied Psycholinguistics*, 35(2), 399–418. <https://doi.org/10.1017/S0142716412000434>
- Bitchener, J., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing*. New York, NY: Routledge.
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., & Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review*, 1(1), 61-67.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-73.

- Bonne, L., & Johnston, M. (2016). Students' beliefs about themselves as mathematics learners. *Thinking Skills and Creativity*, 20, 17–28. <https://doi.org/10.1016/j.tsc.2016.02.001>
- Brantmeier, C., Vanderplank, R., & Strube, M. (2012). What about me? *System*, 40(1), 144–160. <https://doi.org/10.1016/j.system.2012.01.003>
- Broadfoot, P. M., Daugherty, R., Gardner, J., Gipps, C. V., Harlen, W., James, M., et al. (1999). *Assessment for learning: Beyond the black box*. Cambridge, UK: University of Cambridge School of Education.
- Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, 15(2), 119–157.
- Buck, G., Tatsuoka, K. & Kostin, I. (1997). The sub-skills of reading: rule space analysis of a multiple-choice test of second language reading comprehension. *Language Learning*, 47(3), 423–466.
- Burnaby, B. (2008). Language policy and education in Canada. In S. May & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.), Vol. 1: *Language policy and political issues in education* (pp. 331-341). New York: Springer.
- Burner, T. (2014). The potential formative benefits of portfolio assessment in second and foreign language writing contexts: A review of the literature. *Studies in Educational Evaluation*, 43, 139–149. <https://doi.org/10.1016/j.stueduc.2014.03.002>
- Butler, R. (1992). What young people want to know when: Effects of mastery and ability goals on interest in different kinds of social comparisons. *Journal of Personality and Social Psychology*, 62(6), 934-943.
- Butler, R. (1993). Effects of task- and ego-achievement goals on information-seeking during task engagement. *Journal of Personality and Social Psychology*, 65, 18–31.
- Button, S. B., Mathieu, J. E., & Zajac, D. M. (1996). Goal orientation in organizational research: A conceptual and empirical foundation. *Organizational Behavior and Human Decision Processes*, 67(1), 26-48.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219-233.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.

Caruth, G. D. (2013). Demystifying Mixed Methods Research Design: A review of the literature. *Melvana International Journal of Education*, 3(2), 112-122.

Paragon Testing Enterprises. (2017). CELPIP. <https://www.celpip.ca/>

Centre for Canadian Language Benchmarks, The. (2012). *Canadian Language Benchmarks: English as a Second Language for Adults*. Ottawa: Centre for Canadian Language Benchmarks. Accessed at <http://www.cic.gc.ca/english/pdf/pub/language-benchmarks.pdf>

Cervatiuc, A. (2009). Identity, good language learning, and adult immigrants in Canada. *Journal of Language, Identity, and Education*, 8(4), 254-271.

Chalhoub-Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, 19, 273-299.

Chang, H. H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, 80(1), 1-20. <https://doi.org/10.1007/s11336-014-9401-5>

Chanock, K. (2000). Comments on essays: Do students understand what tutors write? *Teaching in Higher Education*, 5(1), 95-105.

Chapelle, C.A., & Chung, Y. (2010). The promise of NLP and speech processing technologies in language assessment. *Language Testing*, 27(3), 301-315.

Chen, H., & Chen, J. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly*, 13(3), 218-230. <https://doi.org/10.1080/15434303.2016.1210610>

Chen, I. J., & Yen, J. C. (2013). Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages. *Computers & Education*, 63, 416-423.

- Chen, Y., Gorin, J.S., Thompson, M.S., & Tatsuoka, K.K. (2008). An alternative examination of Chinese Taipei mathematics assessment: Application of the rule-space method to TIMSS 1999 data. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*. Princeton: IERI. Retrieved from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_01_Chapter_2.pdf
- Chiviawowsky, S., & Wulf, G. (2002). Self-controlled feedback: does it enhance learning because performers get feedback when they need it? *Research Quarterly for Exercise and Sport*, 73(4), 408-418.
- Citizenship and Immigration Canada. (2015). *Language test equivalency charts*. Accessed at <http://www.cic.gc.ca/english/resources/tools/language/charts.asp>
- Clariana, R. B., Wagner, D., & Roher Murphy, L., C. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development*, 48(3), 5-22.
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205-249.
- Clément, R., Baker, S. C., & MacIntyre, P. D. (2003). Willingness to communicate in a second language. *Journal of Language and Social Psychology*, 22(2), 190–209.
<https://doi.org/10.1177/0261927X03022002003>
- Comer, W. J. (2012). Lexical inferencing in reading L2 Russian. *Reading in a Foreign Language*, 24(2), 209-230.
- Corbalan, G., Paas, F., & Cuypers, H. (2010). Computer-based feedback in linear algebra: Effects on transfer performance and motivation. *Computers & Education*, 55(2), 692-703.
- Corpus, J. H., McClintic-Gilbert, M. S., & Hayenga, A. O. (2009). Within-year changes in children's intrinsic and extrinsic motivational orientations: Contextual predictors and academic outcomes. *Contemporary Educational Psychology*, 34(2), 154-166.
<https://doi.org/10.1016/j.cedpsych.2009.01.001>
- Cramp, A. (2011). Developing first-year engagement with written feedback. *Active Learning in Higher Education*, 12(2), 113-124.

Creswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson Education, Inc.

Creswell, J.W., & Plano Clark, V.L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks: Sage Publication.

Crisp, V., & Ward, C. (2008). The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: The PePCAA project. *Computers & Education*, 50(4), 1509–1526. <https://doi.org/10.1016/j.compedu.2007.02.004>

Csizér, K., & Dörnyei, Z. (2005). The internal structure of language learning motivation and its relationship with language choice and learning effort. *Modern Language Journal*, 89(1), 19-36.

Cumming, A., & Berwick, R. (Eds.) (1996). *Validation in language testing*. Philadelphia: Multilingual Matters.

Davis, J. L., Burnette, J. L., Allison, S. T., & Stone, H. (2010). Against the odds: academic underdogs benefit from incremental theories. *Social Psychology of Education*, 41, 331-346. <https://doi.org/10.1007/s11218-010-9147-6>

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, 45(4), 343.

DeMars, C. (2010). *Item response theory*. Oxford: Oxford University Press.

Dempsey, N. P. (2010). Stimulated recall interviews in ethnography. *Qualitative Sociology*, 33(3), 349-367. <https://doi.org/10.1007/s11133-010-9157-x>

DePasque, S., & Tricoli, E. (2015). Effects of intrinsic motivation on feedback processing during learning. *NeuroImage*, 119, 175-186. <https://doi.org/10.1016/j.neuroimage.2015.06.046>

Derwing, T. M., & Waugh, E. (2012). Language skills and the social integration of Canada's adult immigrants. *IRPP Study*, 31.

DiBello, L.V., Stout, W.F., & Roussos, L.A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In Nichols, P.D., Chipman, S.F., & Brennan, R.L. (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Dickinson, D., & Neuman, S. B. (2006). *Handbook of early literacy research: Volume II*. New York, NY: Guildford Press.
- Dihoff, R.E., Brosvic, G.M., Epstein, M.L., & Cook, M.J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record, 54*, 207-231.
- Dikli, S., & Bley, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing, 22*, 1–17.
<https://doi.org/10.1016/j.asw.2014.03.006>
- Dörnyei, Z. (2005) *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9-42). Bristol: Multilingual Matters.
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Oxford: Blackwell.
- Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development, 40*, 315–322. <https://doi.org/10.1016/j.ijedudev.2014.11.004>
- Duchowski, A. (2007). *Eye-tracking methodology: Theory and practice*. London: Springer.
- Duguay, A. L. (2012). ‘The school of life’: Differences in U.S. and Canadian settlement policies and their effect on individual Haitian immigrants’ language learning. *TESOL Quarterly, 46*(2), 306-333.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048.
- Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6*(4), 267–285.

- Educational Testing Service. (2011). *TOEFL® program history. TOEFL iBT® Research Insight, 1(6)*. Princeton: ETS.
- Egi, T. (2008). Investigating stimulated recall as a cognitive measure: Reactivity and verbal reports in SLA research methodology. *Language Awareness, 17(3)*, 212-217.
- Eisenberg, N., Sadovsky, A., & Spinrad, T. L. (2005). Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes. *New Directions for Child and Adolescent Development, 2005(109)*, 109–118. <https://doi.org/10.1002/cd.143>
- El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies, 10(2)*, 121-142.
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*, 218–232.
- Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 70(3)*, 461-475.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28(2)*, 339–368. <https://doi.org/10.1017/S0272263106060141>
- Ertac, S. (2011). Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization, 80(3)*, 532.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272–299.
- Falout, J., Elwood, J., & Hood, M. (2009). Demotivation: Affective states and learning Outcomes. *System: An International Journal of Educational Technology and Applied Linguistics, 37(3)*, 403-417.

- Fernández-Toro, M., & Hurd, S. (2014). A model of factors affecting independent learners' engagement with feedback on language learning tasks. *Distance Education*, 35(1), 106-125. <https://doi.org/10.1080/01587919.2014.891434>
- Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, 32(2), 181-201.
- Field, J. (2008). *Listening in the language classroom*. Cambridge: Cambridge University Press.
- Fink, A. S. (2000). The role of the researcher in the qualitative research process. A potential barrier to archiving qualitative data. *Forum: Qualitative Social Research*, 1(3).
- Fleming, D. (2007). Adult immigrant ESL programs in Canada. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching, Vol. 1* (pp. 185-198). New York: Springer.
- Fontanini, I., & Braga Tomitich, L. M. (2009). Working memory capacity and L2 university students' comprehension of linear texts and hypertexts. *International Journal of English Studies*, 9(2), 1-18.
- Fu, T., & Nassaji, H. (2016). Corrective feedback, learner uptake, and feedback perception in a Chinese as a foreign language classroom. *Studies in Second Language Learning and Teaching*, 6, 161-183.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.
- Furnborough, C., & Truman, M. (2009). Adult beginner distance language learner perceptions and use of assignment feedback. *Distance Education*, 30(3), 399-418. <https://doi.org/10.1080/01587910903236544>
- Garrett, P., & Young, R. F. (2009). Theorizing affect in foreign language learning: An analysis of one learner's responses to a communicative Portuguese course. *Modern Language Journal*, 93(2), 210-226.

- Godfroid, A., & Schmidtke, J. (2013). *What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores*. Honolulu: University of Hawai'i at Manoa, National Foreign Language Resource Center.
- Godfroid, A., & Spino, L. (2015). Reconceptualizing reactivity of think-alouds and eye tracking: Absence of evidence is not evidence of absence. *Language Learning*, 65(4), 896-928
- Goodman, J. S., Wood, R. E., & Chen, Z. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behaviour and Human Decision Processes*, 115(2), 253-267.
- Gorges, J., Kandler, C., & Bohner, G. (2012). Internationalization at home: Using learning motivation to predict students' attitudes toward teaching in a foreign language. *International Journal of Educational Research*, 53, 107–116. <https://doi.org/10.1016/j.ijer.2012.03.001>
- Goto Butler, Y., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal*, 90(4), 506–518.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Gu, M. M., & Cheung, D. S. (2016). Ideal L2 self, acculturation, and Chinese language learning among South Asian students in Hong Kong: A structural equation modelling analysis. *System*, 57, 14–24. <https://doi.org/10.1016/j.system.2016.01.001>
- Haastrup, K. (1991). *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tübingen: Gunter Narr.
- Hailey, D. E., & Hailey, C. (1998). Hypermedia, multimedia, and reader cognition: an empirical study. *Technical Communication*, 45(3), 330–342.
- Han, H. (2009). Institutionalized Inclusion: A case study on support for immigrants in English learning. *TESOL Quarterly*, 43(4), 643–668.

- Harding, L., Alderson, J. C., & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing*, 32(3), 317–336. <https://doi.org/10.1177/0265532214564505>
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hawkey, R., & Milanovic, M. (2013). *Cambridge English Exams – the first hundred years: A history of English language assessment from the University of Cambridge 1913-2013*. Cambridge: Cambridge University Press.
- Higgins, R. (2000). “Be more critical!”: Rethinking assessment feedback. Paper presented at the British Educational Research Association Conference, Cardiff University, Wales.
- Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education*, 6(2), 269-274.
- Hill, S. (2011). Towards ecologically valid assessment in early literacy. *Early Child Development and Care*, 181(2), 165–180. <https://doi.org/10.1080/03004430.2011.536638>
- Hirschel, R., Yamamoto, C., & Lee, P. (2012). Video self-assessment for language learners. *Studies in Self-Access Learning Journal*, 3(3), 291–309.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye-tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- Huang, C. (2011). Achievement goals and achievement emotions: A meta-analysis. *Educational Psychology Review*, 23(3), 359–388. <https://doi.org/10.1007/s10648-011-9155-x>

- Huang, S. C. (2016). Understanding learners' self-assessment and self-feedback on their foreign language speaking performance. *Assessment & Evaluation in Higher Education*, 41(6), 803–820. <https://doi.org/10.1080/02602938.2015.1042426>
- Huff, K.L., & Sireci, S.G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*, 20(3), 16-25.
- Hwang, W.-Y., Shih, T. K., Ma, Z.-H., Shadiev, R., & Chen, S. Y. (2016). Evaluating listening and speaking skills in a mobile game-based learning environment with situational contexts. *Computer Assisted Language Learning*, 29(4), 639–657. <https://doi.org/10.1080/09588221.2015.1016438>
- IELTS. (2017). Common European Framework. Accessed at <https://www.ielts.org/ielts-for-organisations/common-european-framework>
- Jang, E.E. (2005). *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL*. (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Illinois, USA.
- Jang, E.E. (2009). Demystifying a Q-matrix for making diagnostic inferences about L2 reading skills. *Language Assessment Quarterly*, 6, 210-239.
- Jang, E.E. (2010). Demystifying a Q-matrix for making diagnostic inferences about L2 reading skills: The author responds. *Language Assessment Quarterly*, 7, 116-117.
- Jang, E. E., Dunlop, M., Park, G., & Boom, E. H. van der. (2015). How do young students with different profiles of reading skill mastery, perceived ability, and goal orientation respond to holistic diagnostic feedback? *Language Testing*, 32(3), 359-383. <https://doi.org/10.1177/0265532215570924>
- Jang, E. E., Dunlop, M., Wagner, M., Kim, Y. H., & Gu, Z. (2013). Elementary school ELLs' reading skill profiles using cognitive diagnosis modeling: Roles of length of residence and home language environment. *Language Learning*, 63(3), 400–436.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4), 441-480.

- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354.
- Kalyuga, S. (2009). *Managing cognitive load in adaptive multimedia*. Hershey, PA: IGI Publishing.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23-31.
- Kalyuga, S., Chandler, P. and Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, 40, 1-17.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93(3), 579-588.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.
- Kane, M. T. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448-457.
- Kang, H. (2014). Understanding online reading through the eyes of first and second language readers: An exploratory study. *Computers & Education*, 73, 1-8.
<https://doi.org/10.1016/j.compedu.2013.12.005>
- Ke, C. (2006). A model of formative task-based language assessment for Chinese as a foreign language. *Language Assessment Quarterly*, 3(2), 207-227.
https://doi.org/10.1207/s15434311laq0302_6
- Ketabi, S., & Ketabi, S. (2014). Classroom and formative assessment in second/foreign language teaching and learning. *Theory and Practice in Language Studies*, 4(2), 435-440.
<https://doi.org/10.4304/tpls.4.2.435-440>
- Kim, A. Y. (2015). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, 32(2), 227-258.
<https://doi.org/10.1177/0265532214558457>

- Kinlaw, C. R., & Kurtz-Costes, B. (2007). Children's theories of intelligence: Beliefs, goals, and motivation in the elementary years. *The Journal of General Psychology, 134*(3), 295–311.
<https://doi.org/10.3200/GENP.134.3.295-312>
- Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing—An ERP study. *Biological Psychology, 80*(1), 75–83.
<https://doi.org/10.1016/j.biopsycho.2008.03.004>
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.
- Koda, K. (Ed.). (2007). *Reading and language learning*. Malden, MA: Blackwell.
- Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computer-supported, case-based worked examples: Effects of erroneous examples and feedback. *Medical Education, 42*(8), 823-829.
- Kormos, J., Kiddle, T., & Csizér, K. (2011). Systems of goals, attitudes, and self-related beliefs in second-language-learning motivation. *Applied Linguistics, 32*(5), 495–516.
<https://doi.org/10.1093/applin/amr019>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.
- Ku, K. Y. L., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning, 5*(3), 251–267. <https://doi.org/10.1007/s11409-010-9060-6>
- Lam, W. Y. K. (2008). Metacognitive Strategy Use: Accessing ESL learners' inner voices via stimulated recall. *Innovation in Language Learning and Teaching, 2*(3), 207–223.
<https://doi.org/10.1080/17501220802158917>
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics, 1*, 49–74.

- Lau, K., & Chan, D. W. (2007). The effects of cognitive strategy instruction on Chinese reading comprehension among Hong Kong low achieving students. *Reading and Writing, 20*(8), 833–857. <https://doi.org/10.1007/s11145-006-9047-5>
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lea, M. & Street, B. (2000) Student writing and staff feedback in higher education: an academic literacies approach. In M. Lea & B. Stierer (Eds.). *Student writing in higher education: new contexts*. Buckingham: Open University Press.
- Lee, H. C. (2013). Thinking matters: Inferencing in ESL reading lessons. *TESOL Journal, 4*(4), 717-742.
- Lee, C., Cheung, W. K. W., Wong, K. C. K., & Lee, F. S. L. (2013). Immediate web-based essay critiquing system feedback and teacher follow-up feedback on young second language learners' writings: an experimental study in a Hong Kong secondary school. *Computer Assisted Language Learning, 26*(1), 39–60.
- Lee, H. W., Lim, K. Y., & Grabowski, B. L. (2010). Improving self-regulation, learning strategy use, and achievement with metacognitive feedback. *Educational Technology Research and Development, 58*(6), 629-648.
- Lee, M. J., & Tedder, M. C. (2003). The effects of three different computer texts on readers' recall: based on working memory capacity. *Computers in Human Behavior, 19*(6), 767–783.
- Lee, Y. & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6*, 239-263.
- Lee-Thompson, L. C. (2008). An investigation of reading strategies applied by American Learners of Chinese as a foreign language. *Foreign Language Annals, 41*(4), 702-721.
- Leighton, J.P., & Gierl, M.J. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press.
- Leung, C. (2007). Dynamic assessment: Assessment for and as teaching? *Language Assessment Quarterly, 4*(3), 257–278. <https://doi.org/10.1080/15434300701481127>

- Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 9, 17-46.
- Li, H., Hunter, C. V., & Lei, P.W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, 33(3), 391–409.
<https://doi.org/10.1177/0265532215590848>
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309–365. <https://doi.org/10.1111/j.1467-9922.2010.00561.x>
- Li, S. (2014). The interface between feedback type, L2 proficiency, and the nature of the linguistic target. *Language Teaching Research*, 18(3), 373–396.
<https://doi.org/10.1177/1362168813510384>
- Lin, D. M. (2003). Hypertext for the aged: effects of text topologies. *Computers in Human Behavior*, 19(2), 201–209.
- Lin, L. C., & Yu, W. Y. (2015). A think-aloud study of strategy use by EFL college readers reading Chinese and English texts. *Journal of Research in Reading*, 38(3), 286–306.
<https://doi.org/10.1111/1467-9817.12012>
- Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology*, 36(5), 596-613.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27(3), 405-430.
- Magno, C. (2010). The role of metacognitive skills in developing critical thinking. *Metacognition and Learning*, 5(2), 137-156. <https://doi.org/10.1007/s11409-010-9054-4>
- Maier, U. (2009). Towards state-mandated testing in Germany: how do teachers assess the pedagogical relevance of performance feedback information? *Assessment in Education: Principles, Policy & Practice*, 16(2), 205-226. <https://doi.org/10.1080/09695940903076030>

- Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social Cognitive and Affective Neuroscience*, 1(2), 75-86.
- Mantou Lou, N., & Noels, K. A. (2016). Changing language mindsets: Implications for goal orientations and responses to failure in and outside the second language classroom. *Contemporary Educational Psychology*, 46, 22–33.
<https://doi.org/10.1016/j.cedpsych.2016.03.004>
- Mason, B.J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us*. Retrieved from <http://dwb.unl.edu/Edit/MB/MasonBruning.html>
- Matsumoto, M. (2009). Persistence in Japanese language study and learners' cultural/linguistic backgrounds. *Australian Review of Applied Linguistics*, 32(2), 10.1-17.
- McLaren, S. V. (2012). Assessment is for learning: Supporting feedback. *International Journal of Technology and Design Education*, 22(2), 227-245.
- McGlohen, M., & Chang, H. H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior Research Methods*, 40(3), 808–821.
<https://doi.org/10.3758/BRM.40.3.808>
- McLaren, S. V. (2012). Assessment is for learning: Supporting feedback. *International Journal of Technology and Design Education*, 22(2), 227-245.
- McNamara, D., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1–43.
- Melby-Lervag, M., & Lervag, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, 140, 409-433.
- Mendelsohn, D. (2006). Learning how to listen using learning strategies. In Uso-Juan, E., & Martinez-Flor, A. (Eds.). *Current trends in the development and teaching of the four language skills (1–15)*. Munich: Mouton de Gruyter.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-104). New York: Macmillan.

Midgley, C., Kaplan, A., Middleton, M., Maehr M.L, Urdan, T., Anderman L.H, . . . Roeser, R. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology*, 23(2), 113-113.

Midgley, C., Maehr, M. L., Huda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales*. Ann Arbor, MI: University of Michigan. Accessed at http://www.umich.edu/~pals/PALS%202000_V13Word97.pdf.

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered?: it depends on your beliefs about intelligence. *Psychological Science*, 22(3), 320-324.

Miller, T. (2009). Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, 34(2), 181-192.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6–20.
<https://doi.org/10.1111/j.1745-3992.2006.00075.x>

Monopoli, W. J., & Kingston, S. (2012). The relationships among language ability, emotion regulation and social competence in second-grade students. *International Journal of Behavioral Development*, 36(5), 398–405. <https://doi.org/10.1177/0165025412446394>

Montgomery, J.L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, 16, 82-99.

Murphy, L., & Roca de Larios, J. (2010). Feedback in second language writing: An introduction. *International Journal of English Studies*, 10(2), i-xv.

Murphy, P. (2010). Web-based collaborative reading exercises for learners in remote Locations: The effects of computer-mediated feedback and interaction via computer-mediated communication. *ReCALL: The Journal of EUROCALL*, 22(2), 112-134.

- Musso, M., & Cascallar, E. (2009). New approaches for improved quality in educational assessments: Using automated predictive systems in reading and mathematics. *Problems of Education in the 21st Century*, 17, 134–151.
- Nakata, T. (2015). Effects of feedback timing on second language vocabulary learning: Does delaying feedback increase learning? *Language Teaching Research*, 19(4), 416–434.
<https://doi.org/10.1177/1362168814541721>
- Nakayama, A., Heffernan, N., Matsumoto, H., & Hiromori, T. (2012). The influence of goal orientation, past language studies, overseas experiences, and gender differences on Japanese EFL learners' beliefs, anxiety, and behaviors. *Apples – Journal of Applied Language Studies*, 6(2), 19-39.
- Narciss, S. (2004). The impact of informative tutoring feedback and self-efficacy on motivation and achievement in concept learning. *Experimental Psychology*, 51(3), 214-228.
- Nassaji, H. (2016). Anniversary article: Interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research*, 20(4), 535–562. <https://doi.org/10.1177/1362168816644940>
- Neuman, S. B., & Dickinson, D. (2001). *Handbook of early literacy research*. New York: Guildford Press
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Nichols, P. D. (1994). A framework for developing cognitively diagnostic assessments. *Review of Educational Research*, 64(4), 575–603.
- Nosratinia, M., Ghavidel, S., & Zaker, A. (2015). Teaching metacognitive strategies through Anderson's model: Does it affect EFL learners' listening comprehension? *Theory and Practice in Language Studies*, 5(6), 1233-1243.
- Opitz, B., Ferdinand, N. K., & Mecklinger, A. (2011). Timing matters: the impact of immediate and delayed feedback on artificial language learning. *Frontiers in Human Neuroscience*, 5, 8-17.

- Orsmond, P., Merry, S., & Reiling, K. (2005). Biology students' utilization of tutors' formative feedback: A qualitative interview study. *Assessment & Evaluation in Higher Education*, 30(4), 369-386.
- Owen, N. (2016). *An evidence-centred approach to Reverse Engineering: Comparative analysis of IELTS and TOEFL iBT reading sections*. Unpublished doctoral thesis. Leicester: University of Leicester.
- Pappachan, S. (2008). *The influence of goal orientation on the supervisor-subordinate dyad, organizational outcomes, and the feedback environment*. (Published dissertation.)
- Parkison, P. (2009). Political economy and the NCLB regime: Accountability, standards, and high-stakes testing. *The Educational Forum*, 73(1), 44-57.
- Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing*, 15(2), 68-85.
- Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, 92(1), 128–150.
<https://doi.org/10.1037/0021-9010.92.1.128>
- Pitts, S. E. (2005). Testing, testing...: How do students use written feedback? *Active Learning in Higher Education*, 6(3), 218-229.
- Penning de Vries, B., Cucchiarini, C., Bodnar, S., & Strik, H. (2015). Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning: An International Journal*, 28(6), 550.
- Peterson, M. (2016). The use of massively multiplayer online role-playing games in CALL: an analysis of research. *Computer Assisted Language Learning*, 29(7), 1181–1194.
<https://doi.org/10.1080/09588221.2016.1197949>
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26-56.
- Pilipovic, V., & Glusac, T. (2016). Age-related decline in the use of affective learning strategies. *British and American Studies*, 22, 233-252.

- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385–407.
<https://doi.org/10.1007/s10648-004-0006-x>
- Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *Modern Language Journal*, 91(3), 323-340.
- Pokorny, H., & Pickford, P. (2010). Complexity, cues and relationships: Student perceptions of feedback. *Active Learning in Higher Education*, 11(1), 21-30.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3-25.
- Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education*, 33(2), 143-154.
- Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, 36(8), 879-896.
- Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research*, 45(3), 269–286.
<https://doi.org/10.1080/00313830120074206>
- Raofi, S., Chan, S. H., Mukundan, J., & Rashid, S. M. (2014). Metacognition and second/foreign language learning. *English Language Teaching*, 7(1), 36-49.
- Rae, A.M., & Cochrane, D.K. (2008). Listening to students. *Active Learning in Higher Education*, 9(3), 217-230.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372.
- Reb, J., & Connolly, T. (2009). Myopic regret avoidance: Feedback avoidance and learning in repeated decision making. *Organizational Behavior and Human Decision Processes*, 109(2), 182–189.
- Recker, M. M., & Pirolli, P. (1995). Modeling individual differences in students' learning strategies. *Journal of the Learning Sciences*, 4(1), 1–38.

Richardson, D.C., Dale, R., Spivey, M.J. (2006). *Eye movements in language and cognition: A brief introduction*. Retrieved from

http://www.eyethink.org/resources/lab_papers/Richardson2006_Eye_movements_.pdf

Richardson, J. C., Maeda, Y., & Swan, K. (2010). Adding a web-based perspective to the self-assessment of knowledge: Compelling reasons to utilize affective measures of learning. *Academy of Management Learning & Education*, 9(2), 329–334.

Robinson, P. (2013). Aptitude in second language acquisition. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell. DOI: 10.1002/9781405198431.wbeal0035

Rose, H., & Harbon, L. (2013). Self-regulation in second language learning: An investigation of the kanji-learning task. *Foreign Language Annals*, 46(1), 96-107.

Ross, S. (1998). Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1–20.

<https://doi.org/10.1177/026553229801500101>

Roussos, L.A., DiBello, L.V., Stout, W., Hartz, S.M., Henson, R.A., & Templin, J.L. (2007). The fusion model skills diagnosis system. In J.P. Leighton & M.J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 275-318). Cambridge, UK: Cambridge University Press.

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4), 219-262.

Rupp, A. A., Templin, J., & Henson, R. J. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: Guilford Press.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-44.

Saengpakdeejit, R., & Intaraprasert, C. (2014). Reading strategies in foreign language academic reading: A qualitative investigation. *Theory and Practice in Language Studies*, 4(12), 2599-2613.

- Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing*, 25(1), 63–83. <https://doi.org/10.1177/0265532207083745>
- Sawaki, Y., Kim, H., & Gentile, C. (2009). Q-matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments. *Language Assessment Quarterly*, 6(3), 190-209.
- Schlosser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the ‘signal extraction’ counterexplanation for the Dunning-Kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, 39, 85-100. <https://doi.org/10.1016/j.joep.2013.07.004>
- Scott, C. L. (2015). The futures of learning: What kind of learning for the 21st century? *Education Research and Foresight Working Papers*, 14. Paris: UNESCO.
- Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second language proficiency. *Memory & Cognition*, 33(4), 644-653.
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners’ acquisition of articles. *TESOL Quarterly*, 41(2), 255-283.
- Shin, E. C., Schallert, D. L., & Savenye, W. C. (1994). Effects of learner control, advisement, and prior knowledge on young students learning in a hypertext environment. *Educational Technology Research and Development*, 42(1), 33–46.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, 9(2), 169–191.
- Smither, J. W., Brett, J. F., & Atwater, L. E. (2008). What do leaders recall about their multisource feedback? *Journal of Leadership & Organizational Studies*, 14(3), 202-218.

Snow, C. E., (2014). Language, literacy, and the needs of the multilingual child. *Perspectives in Education*, 32(1), 7-16.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.) (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263–308. <https://doi.org/10.1111/j.1467-9922.2010.00562.x>

Statistics Canada. (2009). *The dynamics of overqualification: Canada's underemployed university graduates*. Accessed at <http://www.statcan.gc.ca/pub/11-621-m/11-621-m2006039-eng.htm>.

Statistics Canada. (2012). *Linguistic characteristics of Canadians: Language, 2011 census*. Ottawa: Ministry of Industry. Retrieved October 20, 2014 from <http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-314-x/98-314-x2011001-eng.pdf>

Stump, G. S., Husman, J., & Corby, M. (2014). Engineering students' intelligence beliefs and learning. *Journal of Engineering Education*, 103(3), 369–387. <https://doi.org/10.1002/jee.20051>

Sun, L. (2013). The effect of meta-cognitive learning strategies on English learning. *Theory and Practice in Language Studies*, 3(11), 2004-2009. <https://doi.org/10.4304/tpls.3.11.2004-2009>

Suvorov, R. (2015). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing* 32(4), 463-483. <https://doi.org/10.1177/0265532214562099>

Suzuki, Y. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing*, 32(1), 63–81. <https://doi.org/10.1177/0265532214541885>

Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching*, 46(2), 195-207. <https://doi.org/10.1017/S0261444811000486>

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the Rule Space Method*. New York: Routledge.
- Thompson, T., & Musket, S. (2005). Does priming for mastery goals improve the performance of students with an entity view of ability? *The British Journal of Educational Psychology*, 75(3), 391-409.
- Timmers, C. & Veldkamp, B. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education*, 56, 923-930.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Toulmin, S. E. (2003). *The uses of argument: Updated edition*. Cambridge: Cambridge University Press.
- Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, 19(1), 122-140.
- Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17(4), 292-305.
- Tuckey, M., Brewer, N., & Williamson, P. (2002). The influence of motives and goal orientation on feedback seeking. *Journal of Occupational and Organizational Psychology*, 75, 195.
- UNESCO-UIS. (2016). *Understanding what works in oral reading assessments*. Montreal, QC: UNESCO Institute for Statistics.
- Ushioda, E. (2011). Language learning motivation, self, and identity: Current theoretical perspectives. *Computer Assisted Language Learning: An International Journal*, 24(3), 199-210.
- Uusberg, A., Uiibo, H., Kreegipuu, K., & Tamm, M. (2013). Unintentionality of affective attention across visual processing stages. *Frontiers in Psychology*, 4, 969-970.
- van Beuningen, C. G. de Jong, N. H. & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62(1), 1-41.
- van der Kleij, F.,M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263-272.

- VandeWalle, D. (2003). A goal orientation model of feedback-seeking behavior. *Human Resource Management Review*, 13(4), 581-604.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287-307.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, 23(3), 34-41.
- Wadlinger, H. A., & Isaacowitz, D. M. (2011). Fixing our focus: Training attention to regulate emotion. *Personality and Social Psychology Review*, 15(1), 75-43.
- Wagner, E. (2008). Video listening tests: What are they measuring? *Language Assessment Quarterly*, 5, 218-243.
- Wagner, M. (2015). *The centrality of cognitively diagnostic assessment for advancing secondary school ESL students' writing: A mixed methods study*. Unpublished doctoral dissertation. Toronto: University of Toronto.
- Wagner, M., Chen, M., Park, G., Stone, J., & Jang, E. (2015). Evaluating and revising a reading test blueprint through multiple methods: The case of the CELPIP-General test. Conference presentation at *Language Testing Research Colloquium 2015*, Toronto.
- Walker, M. (2009). An investigation into written comments on assignments: Do students find them usable? *Assessment & Evaluation in Higher Education*, 34(1), 67-78.
- Wang, T. H. (2007). What strategies are effective for formative assessment in an e-learning environment? *Journal of Computer Assisted Learning*, 23(3), 171-186.
<https://doi.org/10.1111/j.1365-2729.2006.00211.x>
- Weaver, M. R. (2006). Do students value feedback? student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379-394.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

- Wildsmith-Cromarty, R., & Conduah, A. N. (2015). 'Push' and 'pull' factors influencing the learning of destination languages by immigrants. *South African Journal of African Languages*, 35(2), 147-155.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education Principles Policy and Practice*, 11(1), 49-65.
- Winke, P., Sydorenko, T., & Gass, S. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *Modern Language Journal*, 97(1), 254-275.
- Yan, D., Almond, R., & Mislevy, R. (2004). *A comparison of two models for cognitive diagnosis*. Educational Testing Service Research and Development Report. Princeton: ETS.
- Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist*, 47(4), 302-314.
- Yeung, A., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology*, 23, 1-21.
- Yoon, E., & Lee, H. K. (2013). Do effects of self-assessment differ by L2 language level? A case of Korean learners of English. *The Asia-Pacific Education Researcher*, 22(4), 731-739.
<https://doi.org/10.1007/s40299-013-0111-z>
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.
- Yorke, M. (2011). Assessment and feedback in the first year: The professional and the personal. In J. Thomas (Ed.), *14th Pacific Rim First Year in Higher Education Conference, Fremantle, Australia, June-July 2011*. Sydney: Colourwise Reproductions.

Zhang, L., Goh, C. C. M., & Kunnan, A. J. (2014). Analysis of test takers' metacognitive and cognitive strategy use and EFL reading test performance: A multi-sample SEM approach. *Language Assessment Quarterly*, 11(1), 76-102.

Zhang, L. J., Gu, P. Y., & Hu, G. (2008). A cognitive perspective on Singaporean primary school pupils' use of reading strategies in learning to read in English. *British Journal of Educational Psychology*, 78(2), 245–271. <https://doi.org/10.1348/000709907X218179>

Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, 15, 3-17.

Zhao, Q., Zhang, J., & Vance, K. (2013). Motivated or paralyzed? Individuals' beliefs about intelligence influence performance outcome of expecting rapid feedback. *Learning and Individual Differences*, 23, 168–171. <https://doi.org/10.1016/j.lindif.2012.07.019>

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70.

Appendices

Appendix 1 Background survey

*Formatting adjusted to fit on page

ESL language learning survey*

What is your name? (Write your full name)

What is your gender? Female Male Wish to specify:

What is your year of birth?

What is your first language? (Only write one)

What languages do you speak at home? (Write all that apply)

Which languages do you know well enough to: (Write all that apply)

Talk with friends

Read a novel

Write a letter or email

Write a work report or college paper

What do you expect to be able to do by the end of your English studies?

	I can already do this	I am sure I will do this	I hope to do this	This will be too difficult for me
Pass English exams (e.g., CELPIP, IELTS, TOEFL)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Get Canadian citizenship	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Get permanent residence in Canada	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Make English-speaking friends	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Read novels in English	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Watch television/video in English	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Work in a job that uses only English	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What is the MAIN reason you are taking English classes? (Only choose one):

<input type="checkbox"/> To get Canadian citizenship	<input type="checkbox"/> To communicate outside the home
<input type="checkbox"/> To get PR (permanent residence) in Canada	<input type="checkbox"/> To talk with family
<input type="checkbox"/> To get a job	<input type="checkbox"/> To make friends
<input type="checkbox"/> Other – Please specify:	

How many years of your life have you spent in an English-speaking country?

How many years of your life have you studied English?

What is the highest level of education have you completed? (Only choose one):

- | | |
|---|--|
| <input type="checkbox"/> None | <input type="checkbox"/> Post-Secondary – one or two years |
| <input type="checkbox"/> Elementary/Primary | <input type="checkbox"/> Undergraduate – three or four year degree |
| <input type="checkbox"/> High school/Secondary school | <input type="checkbox"/> Graduate/Post – graduate degree |

Have you studied in Canada at any of the following levels for your education? (Choose all that apply):

- | | |
|---|--|
| <input type="checkbox"/> None | <input type="checkbox"/> Post-Secondary – one or two years |
| <input type="checkbox"/> Elementary/Primary | <input type="checkbox"/> Undergraduate – three or four year degree |
| <input type="checkbox"/> High school/Secondary school | <input type="checkbox"/> Graduate/Post – graduate degree |

What sector did you work/study in before you came to Canada? (Only choose one):

- | | |
|---|---|
| <input type="checkbox"/> Administration | <input type="checkbox"/> Legal Services |
| <input type="checkbox"/> Construction | <input type="checkbox"/> Media |
| <input type="checkbox"/> Design | <input type="checkbox"/> Personal Services |
| <input type="checkbox"/> Entertainment or Arts | <input type="checkbox"/> Repair, Maintenance, or Installation |
| <input type="checkbox"/> Factory Work, Manufacturing, or Assembly | <input type="checkbox"/> Retail |
| <input type="checkbox"/> Finance or Banking | <input type="checkbox"/> Scientific or Technical |
| <input type="checkbox"/> Fishing, Forestry, Farming, Mining | <input type="checkbox"/> Teaching or Education |
| <input type="checkbox"/> Health or Social Services | <input type="checkbox"/> Transport |
| <input type="checkbox"/> Hotel, Tourism, Restaurant, or Leisure | <input type="checkbox"/> Wholesale |
| <input type="checkbox"/> Worker in the home | |
| <input type="checkbox"/> Other – Please specify: | |

What was your normal role at work before you came to Canada? (Only choose one):

- | | |
|--|---|
| <input type="checkbox"/> Employer | <input type="checkbox"/> Retired |
| <input type="checkbox"/> Senior Employee | <input type="checkbox"/> Student |
| <input type="checkbox"/> Middle or Junior Level Employee | <input type="checkbox"/> Worker in the home |
| <input type="checkbox"/> Self-Employed | |

Over the last three months, have you done any of the following activities? (Choose all that apply):

- | | |
|--|--|
| <input type="checkbox"/> Look after my family | <input type="checkbox"/> Work at an office |
| <input type="checkbox"/> Study English | <input type="checkbox"/> Work in a factory or mill |
| <input type="checkbox"/> Study at secondary school | <input type="checkbox"/> Work in a store or restaurant |
| <input type="checkbox"/> Study at college | <input type="checkbox"/> Work in construction |
| <input type="checkbox"/> Study at university | <input type="checkbox"/> Other – Please specify: |

Do you use English at least three times a week in any of the following activities? (Choose all that apply):

- | | |
|--|--|
| <input type="checkbox"/> Chat online | <input type="checkbox"/> Talk with family |
| <input type="checkbox"/> Communicate in business meetings | <input type="checkbox"/> Talk with friends |
| <input type="checkbox"/> Shop at the grocery market or other store | <input type="checkbox"/> Watch TV or internet videos |
| <input type="checkbox"/> Talk with co-workers or customers | <input type="checkbox"/> Other – Please specify: |

Do you use English at least three times a week in any of the following activities? (Choose all that apply):

<input type="checkbox"/> Communicate on Facebook or Twitter	<input type="checkbox"/> Write business correspondence
<input type="checkbox"/> Read books	<input type="checkbox"/> Write emails
<input type="checkbox"/> Read newspapers or online news stories	<input type="checkbox"/> Write reports for work
<input type="checkbox"/> Read work reports	<input type="checkbox"/> Write school assignments
<input type="checkbox"/> Other – Please specify:	

Read each idea carefully and choose the circle (○) that shows how true each sentence is for you. There is no right or wrong answer. ‘Tasks’ means activities you do for **learning English**. Tasks can include classwork, homework, assignments, practice exercises, projects and other activities you do for learning English.

	Not at all true	A little true	Somewhat true	Fairly true	Very true
With effort, any adult can learn a language fluently	○	○	○	○	○
I only feel successful if other people tell me I did a task well	○	○	○	○	○
I don't want to look stupid so I choose tasks I can do well	○	○	○	○	○
When I do tasks, I want to be more successful than most other people	○	○	○	○	○
I want to do well because I want to show my ability to my family, friends, or teachers	○	○	○	○	○
If I have trouble doing a task, I don't tell anyone	○	○	○	○	○
People are either good or bad at learning languages	○	○	○	○	○
When other people can do a task, I want them to know it is easy for me	○	○	○	○	○
I do not ask questions if I might look stupid	○	○	○	○	○
Good grades are the most important thing for me	○	○	○	○	○
It's important for me to do better than other people on tasks	○	○	○	○	○
It's important to me that I learn new things when I do tasks	○	○	○	○	○
I make sure other people know when I am successful on a task	○	○	○	○	○
I really like to show other people that I can do tasks	○	○	○	○	○
I stay away from tasks where other people might think I'm not smart	○	○	○	○	○
I really want to understand what I am learning	○	○	○	○	○
I prefer challenging and difficult tasks so I can learn new things	○	○	○	○	○
It's important to me that I improve my skills when I do tasks	○	○	○	○	○

If someone works hard, they can learn any language	<input type="radio"/>				
When I am doing tasks, I try to learn new skills	<input type="radio"/>				
It's important to me that I don't look stupid on tasks	<input type="radio"/>				
I choose challenging tasks even if I might not be very successful	<input type="radio"/>				
When I am doing tasks, I enjoy learning as much as I can	<input type="radio"/>				
People can only learn languages well if they are born with language learning ability	<input type="radio"/>				
I prefer easier tasks that make me look good	<input type="radio"/>				

Appendix 2 Self-assessment tool

Name:

Congratulations on finishing the reading test! Read these ideas carefully and think about what you did on the test. Put a check mark on the circle (○) that shows how often each idea was true for you.

While I was doing the test, I could:	Never	Rarely	Sometimes	Usually	Always
understand information that was not obvious in the texts	○	○	○	○	○
put the information in the texts in different words	○	○	○	○	○
know the relationship between the writer and the reader	○	○	○	○	○
know the mood of each text	○	○	○	○	○
understand the main ideas in each text	○	○	○	○	○
understand what would happen even when the text didn't tell me	○	○	○	○	○
find useful information and ignore unimportant information	○	○	○	○	○
know what was coming next in the texts	○	○	○	○	○
use my knowledge of text types (e.g., email, newspaper, menu) to help me understand	○	○	○	○	○
use other information in the texts to understand words I didn't know	○	○	○	○	○
see what were main ideas and what were details	○	○	○	○	○
understand information that was suggested but not written out in the texts	○	○	○	○	○
read the texts quickly to find the information I wanted	○	○	○	○	○
understand descriptions in the texts	○	○	○	○	○
understand the situation, even when the text didn't clearly tell me	○	○	○	○	○
understand the words I read	○	○	○	○	○
understand the facts in the texts	○	○	○	○	○
understand words and phrases used in informal situations	○	○	○	○	○

Appendix 3 Eye-tracking / interview protocol

Eye-Tracking and Interview Protocol

Part I: Starting the Session

[Before starting the session, record the following information on an audio recorder (keep recording through session until end):]

1. Date
2. Interviewer name
3. Student name
4. Location of interview

- **INTRODUCING THE SESSION**

Thanks for agreeing to participate in this activity. In this session you're going to receive your English reading skills feedback report. As I'm interested in what you look at in the report, we're going to set you up with this eye-tracking equipment before you get your report. After we've checked the equipment is working well, you'll access your report and take as long as you need to think about it. When you're done, we'll use the remainder of the time for you to talk me through what you were thinking as you got your report. We'll aim to spend about one hour, and at the end I'll email you a \$10 Amazon gift certificate. Feel free to ask questions at any point during this activity. Do you have any questions right now?

Part II: Eye Tracking

[Set up eye-tracking equipment and calibrate]

Now we're ready, here is your report. Click next, and then you can look at your report. Take as long as you like, and let me know when you're ready to finish. Remember, don't move your head.

[Deactivate equipment]

Part III: Retrospective Think-Aloud

- **INTRODUCING THE RETROSPECTIVE THINK ALOUD**

Now I'm going to show you your report again. I'd like you to talk me through what parts of the report you looked at first, second, and so on. Use as much detail as possible. When you tell me what you looked at, I'd also like you to tell me about what you were thinking when you were viewing that section. Again, use as much detail as possible. I'm interested in what you looked at and what you thought while you were looking, so anything you have to say will be helpful. For example, you can describe things or you can state options. You can say positive or negative things. You can talk as much as you like. I'll ask you questions to help you talk if you seem unsure about what to say. What you say is really important, so I'm going to run this voice recorder to make sure we don't forget anything. *[Check recorder is still running.]* Do you have any questions?

- **RETROSPECTIVELY THINKING ALOUD**

[Show the student their report again.]

Here is your report again. Start thinking aloud when you're ready.

[If the student is verbalizing freely, allow the student to keep talking, and keep notes of any questions to ask the student. If the student stops verbalizing because s/he is looking at the report and thinking, wait for 5 to 6 seconds then prompt him/her with a question. Prompt the student to talk in greater depth with the following types of questions:]

What were you looking for when you looked at [section A]?

What went through your mind when you were looking at [section A]?

I remember you spent some time on [section A] when you first received your report. What were you thinking about?

Can you tell me a little more about that?

What are you thinking now?

Part IV: Debriefing Interview

Thanks very much for telling me about your thoughts while you were looking at the report. Now I just have a few more questions to help me understand how you understand the report.

[Ask the student any questions you noted down during the retrospective think aloud.]

Can you tell me what the report says about your English reading skills?

What do you plan to do with what you learnt in the report?

How have you used the report so far?

[At the end of the interview, thank the student and email them a gift card. Turn off recorder.]

Part V: Interviewer Notes

[After the student has left, make any necessary notes about:

1. The setting (e.g., interruptions, distractions, etc.)
2. Feelings and reflections on events in the session
3. Anything notable about the session]

Appendix 4 Report survey items

1. How much did you look at each part of the report?

	I did not look at this	A little time	Some time	A lot of time	All my time	What is this?
About your report	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Your Canadian English reading skills: about the skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Your Canadian English reading skills: graphs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suggestions for learning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Plan your learning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. Here are 8 statements about the report information. Choose how true each statement is for you.

Statement	Not true	→	→	→	True	I don't understand
There is so much information I cannot remember everything	<input type="radio"/>					
I looked at and thought about all parts of the report	<input type="radio"/>					
I only thought about the very interesting parts	<input type="radio"/>					
I was looking for more information in the report	<input type="radio"/>					
There is too much information in the report	<input type="radio"/>					
I looked at the different parts of the report and thought about how they connected	<input type="radio"/>					
Some part of the report were more interesting than others	<input type="radio"/>					
The information is very general and I want more details	<input type="radio"/>					

3. Here are 8 statements about you and your report. Choose how true each statement is for you.

Statement	Not true	→	→	→	True	I don't understand
I believe what the report says about my English skills	<input type="radio"/>					
I will do what the report told me to do	<input type="radio"/>					
I am comparing the report with my ideas about my English skills now	<input type="radio"/>					
The report is wrong about my English skills	<input type="radio"/>					
I am thinking about how I can use the report for my English studies	<input type="radio"/>					
I am thinking about how I am studying English now	<input type="radio"/>					
The test results are more correct than my self-assessment	<input type="radio"/>					
I am thinking about my English skills now	<input type="radio"/>					

4. Here are 8 statements about using your report. Choose how true each statement is for you.

Statement	Not true	→	→	→	True	I don't understand
I will continue with my English studies without using this report	<input type="radio"/>					
I want to use the report's suggestions in my English studies	<input type="radio"/>					
I know how to use the report for my English learning	<input type="radio"/>					
I want to use my report for my English learning but I am not sure how	<input type="radio"/>					
I want to use the report feedback in my English studies	<input type="radio"/>					
I need help to start using the information in my report	<input type="radio"/>					
I hope to use my report to change my English study activities for better learning	<input type="radio"/>					
I need someone to tell me how to move forward with my English studies now	<input type="radio"/>					

Will you talk to these people about your report?

	Very unlikely	→	→	→	Very likely
My English teacher(s)	<input type="radio"/>				
My English program classmates	<input type="radio"/>				
My friends and family outside English class	<input type="radio"/>				

What will you probably discuss? Select all that apply

- About how my self-assessment compares to the test results
- How much I agree or disagree with the report
- How to improve my English
- My level of English proficiency
- What areas of my English skills I should focus on
- What the other person thinks about my English skills
- What the report says about my English language skills

Appendix 5 Delayed recall interview protocol

Follow-up interview protocol

Part I: Starting the Session

[Before starting the session, record the following information on an audio recorder (keep recording through session until end):]

1. Date
2. Interviewer name
3. Student name
4. Location of interview

- **INTRODUCING THE SESSION**

Thanks for agreeing to talk with me today. In this session I'm going to ask you to talk to me about what you've been studying, and about your feedback report. We'll plan to spend about 30 minutes, and at the end I'll email you a \$10 Amazon gift certificate. You can ask questions any time you want. Do you have any questions right now?

Part II: Interview

[Interview begins. Work through the following main topics with questions similar to those listed below.]

Current studies

Can you tell me a little bit about what you've been studying this month?

Unstructured recall of usage

How have you used your reading report?

Recall of report content

Tell me what you remember about your reading report.

What did it tell you about your reading skills?

What did the report look like?

What did your report suggest you should do to continue learning?

What learning goals did you set?

Structured recall of usage

I remember you set learning goals X, Y and Z. Can you tell me how you've been working on these goals this month?

Why have you chosen to do these learning activities?

How have you been progressing in your learning goals?

[Elicit deeper recall using questions similar to those listed below]

Can you tell me a little more about that?

What do you mean, XXX?

What are you thinking now?

Why do you think that?

[At the end of the interview, thank the student and email them a gift card. Turn off recorder.]

Part III: Interviewer Notes

[After the student has left, make any necessary notes about:

1. The setting (e.g., interruptions, distractions, etc.)
2. Feelings and reflections on events in the session
3. Anything notable about the session]

Appendix 6 Exploratory factor analyses

Exploratory factor analyses were selected over principal component analyses because it was hypothesized that latent variables were being measured, and the ways that the items correlated to represent these variables was of interest, as opposed to simply desiring to reduce the items as much as possible based on their shared variance (Fabrigar, Wegener, MacCallum, & Strahan, 1999). All the exploratory factor analyses were conducted using principal axis factoring because it can be used when the assumption of normality is violated (Fabrigar, Wegener, MacCallum, & Strahan, 1999), as is consistently the case in this data set. Oblique rotation (promax) was used for all the analyses because the hypothesized factors were assumed to be correlated, and oblique rotation is computationally simpler compared to the other common oblique rotation method (oblimin), yet appears to have no comparative weaknesses, so as the more concise option, it was selected. The number of factors was accepted based on the percentage of variance explained by the factor (e.g., scree plots) to ensure that all factors explaining a substantial amount of variance were included.

Background survey: Goal orientation items

Table 38

Summary of exploratory factor analysis results for goal orientation items

Item	Factor Loadings *		
	Factor 1	Factor 2	Factor 3
m1	It's important to me that I learn new things when I do tasks	.60	
m2	I really want to understand what I am learning	.59	
m3	When I am doing tasks, I try to learn new skills	.62	
m4	I prefer challenging and difficult tasks so I can learn new things	.53	
m5	I choose challenging tasks even if I might not be very successful	.44	
m6	When I am doing tasks, I enjoy learning as much as I can	.76	
m7	It's important to me that I improve my skills when I do tasks	.65	
pp1	When I do tasks, I want to be more successful than most other people	.59	
pp2	I make sure other people know when I am successful on a task	.70	
pp3	When other people can do a task, I want them to know it is easy for me	.41	

pp4	It's important for me to do better than other people on tasks		.69		
pp5	I only feel successful if other people tell me I did a task well		.32		
pp6	I want to do well because I want to show my ability to my family, friends, or teachers		.72		
pp7	I really like to show other people that I can do tasks		.76		
pa1	I do not ask questions if I might look stupid			.52	
pa2	It's important to me that I don't look stupid on tasks			.66	
pa3	I prefer easier tasks that make me look good			.45	
pa4	I don't want to look stupid so I choose tasks I can do well			.68	
pa5	Good grades are the most important thing for me		.53		
pa6	I stay away from tasks where other people might think I'm not smart			.67	
pa7	If I have trouble doing a task, I don't tell anyone			.32	
Eigenvalues			3.75	1.81	1.59
% of variance			32	27	25
Correlations between factors		Factor 2	-.02	1.00	
		Factor 3	.29	-.21	1.00

N = 90, method = principal factors, rotation method = promax. * Factor loadings > .30 displayed.

Report survey: Items for perceptions about amount of information

Table 39

Summary of exploratory factor analysis results for items on perceptions about amount of information: all survey items

Item	Factor Loadings *			
	Factor 1	Factor 2	Factor 3	Factor 4
a1			.65	
a2	.36		.34	
a3	.53			
a4	.58			
b1		.55		
b2				.51

b3	I look at the different parts of the report and through about how they connected				.55
b4	I only thought about the very interesting parts			.55	
Eigenvalues		1.37	0.86	0.39	0.25
% of variance		68	42	19	12
Correlations between factors		Factor 2	.30	1.00	
		Factor 3	.24	.38	1.00
		Factor 4	.43	.98	-.19
					1.00

N = 88, method = principal factors, rotation method = promax. * Factor loadings > .30 displayed.

Table 40

Summary of exploratory factor analysis results for items on perceptions about amount of information: used survey items

Item	Factor Loadings *	
	Factor 1	
a1	There is too much information to understand	.41
a2	There is so much information I cannot remember everything	.54
b4	I only thought about the very interesting parts	.50
Eigenvalue		0.71
% of variance		176

N = 98, method = principal factors, rotation method = promax. * Factor loadings > .30 displayed.

Report survey: Items for perceptions of content of report

Table 41

Summary of exploratory factor analysis results for items on perceptions of content of report

Item	Factor Loadings *		
	Factor 1	Factor 2	Factor 3
a1	I will do what the report told me to do		.61
a2	The test results are more correct than my self-assessment	.57	
a3	I believe what the report says about my English skills	.74	
a4	The report is wrong about my English skills	-.48	
b1	I am comparing the report with my ideas about my English skills now		.59
b2	I am thinking about my English skills now	.59	

Item	Factor Loadings *		
	Factor 1	Factor 2	Factor 3
b3 I am thinking about how I can use the report for my English studies		.60	
b4 I am thinking about how I am studying English now		.74	
Eigenvalues	2.32	0.66	0.50
% of variance	80	23	17
Correlations between factors	Factor 2	.48	1.00
	Factor 3	.33	.30

N = 90, method = principal factors, rotation method = promax. * Factor loadings > .30 displayed.

Report survey: Items for planned usage of report

Table 42

Summary of exploratory factor analysis results for items on planned usage of report

Item	Factor Loadings *		
	Factor 1	Factor 2	
a1 I know how to use the report for my English learning	.33		
a2 I need help to start using the information in my report		.77	
a3 I need someone to tell me how to move forward with my English studies now	.33	.60	
a4 I want to use my report for my English learning but I am not sure how		.63	
b1 I want to use the report feedback in my English studies	.70		
b2 I hope to use my report to change my English study activities for better learning	.65		
b3 I want to use the report's suggestions in my English studies	.72		
b4 I will continue with my English studies without using this report		.35	
Eigenvalues	2.53	1.02	
% of variance	76	30	
Correlations between factors	Factor 2	.20	1.00

N = 92, method = principal factors, rotation method = promax. * Factor loadings > .30 displayed.