Understanding the Causes of Inaccurate Self-Assessments: Extraversion's Role

Caitlin Gaffney

1. Introduction

Self-assessment, also known as self-evaluation or self-rating, is frequently used to measure L2 ability (Suzuki, 2015) both in classrooms to encourage reflection on L2 learning (e.g., Little, 2005; Glover, 2011) and in proficiency tests (e.g., DIALANG; Alderson, 2005) as an alternative to formal L2 assessment for placement and criterion-referenced interpretations (Ross, 1998). However, research on the construct validity of L2 self-assessment is mixed, with some studies demonstrating that the tool provides reliable estimations of L2 ability (e.g., Blanche & Merino, 1989; Brantmeier & Vanderplank, 2008; Luoma, 2012; Préfontaine, 2013), but others finding weak or non-existent relationships with more objective measures of linguistic skills (e.g., Raasch, 1980; Brantmeier, 2006; Foote, 2010; Suzuki, 2015). An incompatibility between real and perceived L2 competence is problematic for many reasons including the fact that self-assessed ability can have behavioural consequences, affecting confidence and willingness to communicate in class (de Saint Léger, 2009; de Saint Léger & Storch, 2009). As a result of flawed self-assessed ability, under-confident learners may avoid participating in L2 communication, and over-confident learners may not take advantage of opportunities to improve their L2 skills when necessary.

We see then that understanding the causes behind inaccurate self-assessment is crucial. That said, in the case of studies that have not found a relationship between self- and objectively assessed L2 proficiency, only a handful of researchers have attempted to explain the mismatch, suggesting several potential causes including amount of L2 experience and feedback (Trofimovich, Isaacs, Kennedy, Saito, & Crowther, 2016), or psychosocial or individual factors (Dlaska & Krekeler, 2008). Even fewer researchers have sought to empirically explain the difference between self- and other-assessed ability (exceptions: MacIntyre, Noels, & Clément, 1997; Trofimovich et al., 2016). We propose here that along with other variables, a personality trait, extraversion, may contribute to explaining inaccurate self-assessments since extraverts are more assertive, enthusiastic (Hirsh & Peterson, 2008), and experience a higher sense of personal accomplishment (Thoresen, Kaplan, Barsky, Warren, & de Chermont, 2003) than their more cautious (Dewaele, 2002), introverted counterparts.

*Caitlin Gaffney, University of Toronto, Caitlin.gaffney@mail.utoronto.ca.

Another factor that has been shown to cause variation in the accuracy of L2 self-assessments is the skill being assessed (e.g., speaking, listening, reading, writing; Butler & Lee, 2010). While studies have examined learners' ability to self-assess speaking (e.g., Dlaska & Krekeler, 2008), listening (e.g., Brantmeier, Vanderplank & Strube, 2012), reading (e.g., Suzuki, 2015), and writing (e.g., Yaghoubi-Notash, 2012), there is a lack of research examining self-assessment's efficacy for the evaluation of vocabulary, an integral source of knowledge for any L2 communication. These lacunae in the research on L2 self-assessment's validity warrant a study that: (1) examines the ability of L2 learners to self-assess their L2 vocabularies; and (2) investigates the possibility that extraversion contributes to inaccurate self-assessments. In order to fulfill these objectives, we answer the following questions: (1) Is there a mismatch between self- and objectively assessed L2 vocabulary?; and (2) To what extent can extraversion explain variability observed in self-assessed L2 vocabulary?

We start with a discussion of research on the validity of L2 self-assessments (§2.1), before presenting studies which have attempted to empirically explain the mismatch between self- and other- assessed L2 ability (MacIntyre et al., 1997; Trofimovich, et al., 2016; §2.2). Then, in Section 2.3, we introduce the independent variable, extraversion, and explain why it may influence learners' self-assessments. In Section 3, we introduce the experimental study, including the hypotheses, participants, and methodology. In Section 4, we present the results, which are discussed in detail in Section 5, before presenting limitations and suggestions for future research in Section 6. We conclude in Section 7.

2. Background and motivation for the current study

The following sections present the relevant literature on L2 self-assessment validity (§2.1), and on variables which explain the inability of L2 learners to accurately assess their competence (§2.2). Lastly, we discuss the potential of extraversion to determine variation in self-assessed L2 vocabulary (§2.3).

2.1. Validity of self-assessments

While L2 self-assessments were previously thought to be good or very good measures of L2 ability (see the following literature review: Blanche & Merino, 1989), a meta-analysis conducted by Ross (1998) demonstrated variation in learners' ability to self-assess their L2 skills. Since then, recent studies have shown low or non-existent correlations between self- and other-assessed skills (e.g., Brantmeier et al., 2012). This is in keeping with psychological and educational research showing that "people's self-views hold only a tenuous to modest relationship with their actual behavior and performance" (Dunning, Heath, & Suls, 2004, p. 69). In the current section, we discuss four studies which have compared learners' self-assessed L2 skills to more objective measures of L2 ability in order to investigate whether or not self-assessment is a valid measure of L2 reading, listening, speaking, and writing.

Brantmeier (2006) investigated the ability of 71 advanced L2 learners of Spanish to self-assess their reading ability. After reading a 1270-word passage, learners completed three in-class reading comprehension tasks (written recall, sentence completion, and multiple choice) then self-assessed their L2 reading using a questionnaire consisting of two 5-point Likert scale questions. Results from a multiple regression analysis revealed that none of the three reading task scores were significant predictors of self-assessed reading ability, suggesting that learners could not accurately estimate their L2 reading skills as measured in class. Brantmeier suggested that the self-assessment instrument may not have been broad and descriptive enough, preventing learners from expressing exactly what they were capable of doing with the L2.

Addressing this methodological limitation, Brantmeier et al. (2012) tested whether 150 advanced classroom learners of L2 Spanish (L1 English) could accurately self-assess L2 skills when using specific criterion-referenced scales. The scales included descriptions of the specific knowledge and skills being examined, and were tailored to the course objectives described in the syllabus and the experience learners had in practicing the skills. L2 skills were objectively measured using an online multiple-choice diagnostic test created to tap into specific concepts taught in the course. Weak, but significant correlations were found between self- and objectively assessed reading ($r = .229$, $p < .01$) and listening ($r = .274$, $p < .01$), demonstrating that even when using scales specifically designed for given learner groups, a misalignment between self-assessed and actual ability remains. The researchers admitted that the weakness of the relationships suggested that other important moderators of self-assessments need to be considered (p. 150).

Although early studies on self-assessment validity first emphasized the receptive skills (listening and reading; Brantmeier et al., 2012), research has now been extended to speaking (e.g., de Saint Léger, 2009) and writing (e.g., Yaghoubi-Notash, 2012). Despite the fact that learners' ability to self-assess reading and listening did not seem to differ greatly in Brantmeier et al. (2012), research has shown that accuracy of self-assessment varies according to the skill being assessed, with weaker relationships being reported in the productive skills (speaking and writing; Ross, 1998). For example, Dlaska and Krekeler (2008) revealed that experienced L2 learners find it difficult to correctly self-assess pronunciation skills. In an investigation of L2 German vowel and consonant production by advanced classroom learners (various L1s), 46 participants produced various consonant and vowel sounds before comparing their phoneme productions to that of a native speaker. Their productions were then self-assessed and objectively assessed by two trained raters. Although learners could effectively identify their L2 sounds as accurate when they were in fact accurate, they could only identify 44% of their inaccurate productions as such. This study underlined the gap between self- and objectively assessed L2 production of phonemes, even when learners were able to directly compare their productions with native speaker phonemes.

Lastly, Yaghoubi-Notash (2012) examined the validity and reliability of self-assessed L2 writing in 64 Iranian learners of L2 English (L1s: Azerbaijani,

Persian, and Kurdish). Learners completed a task that required them to write five paragraphs on pre-selected topics in 60 minutes. Students rated their essays both holistically out of 100 and according to various categories on an essay evaluation sheet. The teacher then assessed the essays in the same two ways (holistic and category-based). An objective measure of writing ability was created by averaging the teacher's holistic and category-based assessments with students' writing grades from the previous semester. Results revealed significant correlations between the objective measure and the holistic ($r = 0.29$, $p < .05$) and category-based self-assessed writing ($r = 0.41$, $p < .01$). Once again, we see that the ability to self-assess L2 skill can vary greatly according to the self-assessment tool in use. Despite the weak and moderate correlations however, the author did not consider self-assessment a valid measure of L2 writing.

We see in the studies described above, that although the strength of the relationship between self- and other-assessed L2 ability seems to vary from study to study, self-assessment is not strongly related to other measures of L2 reading, listening, speaking, and writing. When reviewing the literature on L2 self-assessment validity however, we must take into account the fact that results differ based on various between-study methodological differences (Blanche & Merino, 1989; Ross, 1998; Brantmeier et al., 2012) including the self-assessment instruments employed (e.g., Likert scales, descriptive criterion-referenced measures, relation to curricular content), the types of objective measures of L2 ability (e.g., teacher-assessments, written tests, native-speaker assessments), the amount of L2 experience and/or proficiency of the participants, and the specific linguistic skills being tested (speaking, listening, reading, writing). It is therefore integral that researchers further investigate how the accuracy of L2 assessment varies according to different parameters.

As previously mentioned, vocabulary, a prerequisite for L2 communication, has yet to be studied. Vocabulary may be especially difficult to self-assess for two reasons. First, in "ill-defined, fuzzy, or highly subjective domains, such as intelligence, sophistication, or idealism, [self-assessment is] more inflated than in more specific domains like neatness or punctuality (Dunning, Meyerowitz & Holzberg, 1989; Hayes & Dunning, 1997; Burson, Larrick & Klayman, 2006)" (as cited in Trofimovich et al., 2016, p. 123). It has indeed been found that individuals tend to overestimate their abilities on tests of acculturated knowledge (e.g., vocabulary and general knowledge tests; Juslin, 1994; Stankov, 1998). Second, a lack of feedback has been thought to lead to inaccurate self-assessments (Trofimovich et al., 2016). As feedback regarding vocabulary is not offered within classroom contexts as often as feedback regarding grammatical accuracy and pronunciation, classroom learners may be more out of touch with the reality of their lexical ability.

In this section, we have shown that L2 learners are frequently unable to accurately self-assess their L2 skills. The studies cited to this point highlight the fact that, as suggested by Brantmeier et al. (2012), variables aside from actual L2 competence may be responsible for predicting self-assessed L2 ability. In the following section, we present the little research which has studied other causes behind the observed mismatch between self- and objectively assessed L2 skills.

2.2. Studies explaining inaccurate L2 self-assessments

In the previous section, we observed that a consistent mismatch between actual and self-assessed L2 ability has been found in recent research on self-assessment's construct validity. If there exists misalignment between self-perception and reality, learners may not notice when their L2 performance differs from that of target-like language (Trofimovich et al., 2016). Without noticing how their linguistic ability differs from native speakers', they may behave in ways that are detrimental to their L2 acquisition. Despite the importance of this problem, a question that has received a surprising lack of attention in the L2 assessment literature is *why* learners experience problems perceiving their L2 ability. We previously noted that this could be explained by methodological variables, but several researchers have suggested that individual differences may also act as important mediators in self-assessments (e.g., Butler & Lee, 2010). Among the suggestions for the misalignment are previous experience and skeptical/over-critical attitudes (Dlaska & Krekeler, 2008), social-psychological factors (e.g., gender, social norms, culture) that are linked to overconfidence (Trofimovich et al., 2016), age (Butler & Lee, 2006), the extent to which the students can use episodic memory to exercise the skills that are assessed (Ross, 1998), and learners' self-esteem and motivation (Anderson, 1982; AlFallay, 2004; Dörnyei, 2001). Despite these suggestions, however, a real lack of research using valid and reliable measures of these individual differences has sought to empirically demonstrate their effects on self-assessments. Two exceptions are MacIntyre et al. (1997) and Trofimovich et al. (2016).

MacIntyre et al. (1997) tested whether language anxiety would be a significant predictor of self-assessed L2 ability on 26 different French tasks (eight speaking, seven listening, five reading, and five writing) after controlling for actual task performance. 37 Anglophone learners varying greatly in L2 proficiency completed language anxiety tests, self-ratings of their perceived competence on the 26 French tasks, and then the 26 tasks themselves. All correlations between self-rated competence and language anxiety were negative, ranging from $r = -.52$ for reading to $r = -.60$ for speaking. A regression analysis controlling for actual competence demonstrated that anxious learners underestimated their ability compared to less anxious students who more often overestimated their performance. The study identified language anxiety as an important variable in influencing self-assessments of all L2 skills.

Trofimovich et al. (2016) studied the relationship between self- and rater-assessed accentedness and comprehensibility in L2 English speech (various L1s) in an attempt to find further support for the Dunning-Kruger effect within the field of L2 acquisition. The Dunning-Kruger effect posits that due to a lack of information and awareness, low performers tend to over-estimate their ability when compared to high performers (Dunning et al., 2004; Carter & Dunning, 2008 as cited in Trofimovich et al., 2016). The researchers found that in 132 learners, most self-assessments were inaccurate (no correlation between speakers' actual and self-rated scores for accent $r = .06$, $p = .50$; weak correlation between actual and self-rated scores for comprehensibility $r = .18$, $p = .03$). Furthermore, the

researchers created over-confidence scores for the participants by subtracting the mean rating for each speaker from the speaker's self-rating and expressing the difference as a proportion on a 9-point scale. Overconfidence scores for the learners in the top and bottom thirds were then compared. For both accentedness and comprehensibility, the bottom thirds of the speakers were significantly more overconfident than the top thirds (accentedness: $M = .27$ or $+2.4$ points on a 9-point scale versus $M = -.19$ or -1.7 points, $t(88) = 19.39$, $p < .001$; comprehensibility: $M = .18$ or $+1.6$ points on a 9-point scale versus $M = -.23$ or -2 points, $t(88) = 19.13$, $p < .001$). We see then that learners with the least amount of ability have the most confidence in terms of self-assessments, whereas the learners who have greater ability under-estimate their skill.

These two studies confirm the existence of other important individual differences that lead to variation in L2 self-assessments. We add to this line of research by considering a personality trait's effect. Previous studies seem to suggest, without naming extraversion explicitly, that stable characteristics linked to confidence may be responsible for variation in self-assessments. For example, based on anecdotal evidence from two participants, Dlaska and Krekeler (2008) suggested that over-critical and skeptical attitudes influenced self-assessments. Additionally, Trofimovich et al. (2016) advised that research should consider social-psychological factors linked to over-confidence. Accordingly, we explore the possibility that a personality trait, extraversion, explains variability observed in self-assessed vocabulary.

2.3. Personality/extraversion

Personality is defined as 'consistent patterns in the way individuals behave, feel and think' (Pervin & Cervone, 2010, p. 228). It predicts '…happiness, physical and psychological health, spirituality, and identity…the quality of relationships with peers, family, and romantic others at an interpersonal level…occupational choice, satisfaction, and performance…community involvement...political ideology' (Ozer & Benet-Martínez, 2006, p. 401). It is thus somewhat surprising that research has not yet considered the effect of personality on L2 assessments.

In recent personality research, the Big Five model (e.g., McCrae & Costa, 2003) is almost ubiquitous (Funder, 2001). It adopts the first two dimensions from Eysenck's Personality theory (Eysenck & Eysenck, 1985; among others), extraversion and neuroticism, but replaces the third, psychoticism, with conscientiousness, agreeableness, and openness to experience. Born out of the hypothesis that factor analysis could determine broad universal dimensions based on an exhaustive list of thousands of personality adjectives, the five traits have been found to be effective descriptors of individuals' behavior, and remain stable throughout life (Costa & McCrae, 1997) and across cultures (McCrae & Costa, 2003). Each trait is bi-polar (e.g., extraversion-introversion) and non-categorical, meaning that according to this model, an individual is not either an introvert or an extravert, but they fall somewhere between. Since extraversion seems most likely to influence self-perception, we focus solely on this trait in the current study.

Extraverts are confident, and optimistic (Chamorro-Premuzic & Furnham, 2004), they demonstrate assertiveness and enthusiasm (Hirsh & Peterson, 2008), excitement-seeking, and positive emotions (Dörnyei, 2005), and they experience a higher sense of personal accomplishment (Thoresen et al., 2003). On the other hand, introverts are more cautious (Dewaele, 2002), reserved, restrained (Dörnyei, 2005), and self-critical (McPeek, Nichols, Classen, & Breiner, 2011). We can see then why extraversion may significantly predict variability in self-assessments.

After reviewing the pertinent literature on the validity of L2 self-assessments, and variables that might explain the mismatch between L2 self- and objectively assessed ability, we identify two main lacunae in the research. First, while research has investigated the validity of L2 self-assessments in speaking, writing, reading, and listening, studies have yet to examine the efficacy of this assessment tool in evaluating vocabulary, a prerequisite for all of these skills. Second, although gaps between self- and other-assessed L2 ability have repeatedly been found in the research, there is a dearth of studies empirically linking predictors to variability in inaccurate self-assessments. A study that considers the potential role of personality in influencing the self-perception of L2 vocabulary is therefore warranted.

3. Present study

To address the previously mentioned lacunae, the current study explores extraversion's role in explaining variability observed in self-assessed vocabulary. We make a second contribution to existing literature by investigating learners' ability to accurately self-assess vocabulary, an aspect of L2 ability which has been ignored in previous research, despite its importance for L2 communication.

3.1. Research questions and hypotheses

The present study was designed to answer the following research questions:

(1) Is there a mismatch between self- and objectively assessed L2 vocabulary?

As previously noted, prior work has revealed that self-assessment varies according to the skill being assessed (e.g., Ross, 1998). In comparison to other L2 skills, vocabulary may be more abstract, rendering it more difficult to self-evaluate. In addition, classroom learners likely receive less feedback related to lexical knowledge than in other L2 domains (e.g., speaking, writing), making it harder to self-rate. It could therefore be more difficult for learners to assess their knowledge as they are unaware of what constitutes an "excellent" vocabulary. We therefore expect that learners will not accurately self-assess their L2 vocabulary and expect to see no correlation between self-assessed and objectively measured vocabulary.

(2) To what extent can extraversion explain variability observed in self-assessed L2 vocabulary after accounting for objectively measured L2 vocabulary?

Extraverted learners are more assertive, enthusiastic (Hirsh & Peterson, 2008), and experience a higher sense of personal accomplishment (Thoresen et al., 2003) than more cautious (Dewaele, 2002), self-critical (McPeek et al., 2011) introverted counterparts. We therefore predict that after accounting for objectively measured L2 vocabulary, extraversion will be a significant predictor of self-assessed L2 vocabulary.

3.2. Learners

55 L1 English university learners of L2 French (M_{age} = 20.25, SD = 1.70; 5 males, 50 females) participated. As this data was collected within the context of a larger research project, recruitment was restricted to those enrolled in intermediate-advanced French courses at the time of participation. To avoid burying the effect of extraversion with variation in actual L2 vocabulary ability, an objective measure of lexical proficiency (see §3.3.2) was used and entered into the regression analysis as an independent variable.

3.3. Instruments

Three instruments were used in the study: a personality test to measure extraversion, a brief background questionnaire to obtain biographical data and ask participants to assess their L2 vocabularies, and a lexical proficiency test to tap into L2 vocabulary.

3.3.1. Personality test

Extraversion was operationalized using the valid and reliable Big Five Aspect Scales (BFAS) test (DeYoung, Quilty, & Peterson, 2007) which consists of 100 items (e.g., 'Make friends easily.') rated on a five-point Likert scale (strongly disagree [1] to strongly agree [5]). While the test provides scores for all Big Five personality traits, we focus just on extraversion in this study.

3.3.2. Linguistic background questionnaire and Vocabulary Levels test

Jeffrey Steele and colleagues are currently developing a new online placement test (University of Toronto French Proficiency Test; UTFPT) designed for French-as-a-second-language courses. Participants completed two subcomponents of the UTFPT test: a linguistic background questionnaire, which asked learners to share information regarding current French course, sex, age, L1, and known languages, and a Vocabulary Levels test (VLT) to evaluate vocabulary breadth. Before beginning the VLT, learners were presented with the instructions: 'From the six choices on the left, choose the appropriate word for each definition

on the right.' The 45 items varied according to lexical frequency (based on words' frequency ranking bands; 0–1000; 1001–2000; 2001–3000; 3001–4000; or 4001–5000) and grammatical category. Each prompt included three simple definitions made up of words from the same or lower frequency bands as the responses. The six words (three target responses and three distractors) belong to the same frequency band.

3.3.3. Self-assessments

As part of a linguistic background questionnaire, learners received the following question: "How would you rate your vocabulary? Please check as appropriate." Learners then rated their L2 vocabulary as Poor (1), Fair (2), Good (3), Very good (4), or Excellent (5).

4. Results

Self-assessed L2 vocabulary was first correlated with objectively measured L2 vocabulary (VLT scores), and extraversion to determine the strength and significance of the relationships among the variables. Subsequently, a multiple linear regression analysis with self-assessed L2 vocabulary as the dependent variable was used to test whether extraversion would add in predictive power for L2 assessments after VLT scores were already added to the equation.

4.1. Correlational analyses

The relationships between self-assessed L2 vocabulary and objectively measured vocabulary, and extraversion were tested using Pearson Product Moment Correlations (2-tailed; probability level set at $p < .05$). As shown in Table 1, no correlation exists between L2 self-assessed vocabulary and VLT scores. On the other hand, we see a strong significant correlation between self-assessed L2 vocabulary and extraversion. These findings support results from previous research demonstrating that L2 self-assessments correlate weakly (e.g., Brantmeier et al., 2012; Yaghoubi-Notash, 2012) or not at all (e.g., Brantmeier, 2006; Foote, 2010; Trofimovich et al., 2016) with more objective measures of L2 ability. The strength of the correlation between self-assessed vocabulary and extraversion suggest that the latter will be a significant predictor of variance in self-assessments.

Table 1. Correlations between learners' self-assessed L2 vocabulary and independent variables ($n = 55$)

Dependent variable	Independent variables	r	p
Self-assessed L2 vocabulary	VLT scores	.14	.16
	Extraversion	.39	.00

To address our second research question, we used a regression analysis to examine extraversion's unique effect on L2 self-assessed vocabulary after accounting for actual L2 vocabulary.

4.2. Sequential regression analysis

A sequential regression process, as opposed to a standard or stepwise regression process, was used since previous research has already established relationships between L2 self-assessments and objectively measured ability (e.g., Suzuki, 2015, Trofimovich et al., 2016). By entering VLT scores first, we were able to see if extraversion would explain L2 self-assessed vocabulary after controlling for this factor. In looking at the R^2 contribution of extraversion, we determine how much variance in self-assessed vocabulary is explained independently by this variable. The significance level was again set to 0.05. As seen in Table 2, VLT scores were not found to be a significant predictor of L2 self-assessed vocabulary ($p = .22$) in the regression analysis. Extraversion, however was a significant predictor of L2 self-assessed vocabulary ($p < .05$) accounting for 13% of observed variance.

Table 2. Summary of regression analysis of independent variables predicting self-assessed L2 vocabulary

Model	Total R^2	ΔR^2	VLT β	Extraversion β
1	.02	.02	.02 (-.02, .05)	
2	.16*	.13*	.01 (-.02, .04)	.05* (.02, .09)

* p < .05

5. Discussion

Previous research had found weak or non-existent correlations between self- and objectively assessed L2 speaking (e.g., Dlaska & Krekeler, 2008), listening (e.g., Brantmeier et al., 2012), writing (e.g., Yaghoubi-Notash, 2012), and reading (e.g., Brantmeier, 2006). The finding that the subjective measure of self-assessment lacks validity seems logical since people have been described as "famously bad at judging their own ability" (Trofimovich et al., 2016, p. 122). One limitation in the previous research however, was that it had not examined the validity of self-assessments of L2 vocabulary, despite the fact that learners' ability to self-assess L2 competence varies according to the skill being assessed (e.g. Ross, 1998). We predicted no significant correlation between self- and objectively measured vocabulary size. We expected that in comparison to the other four skills, this type of L2 knowledge may be especially difficult to assess due its abstract nature and the lack of in-class feedback given regarding vocabulary. Our results supported the hypothesized mismatch between self- and objectively measured vocabulary and no significant correlation was found between VLT scores and self-

assessed L2 vocabulary ($r = .14$, $p = .16$). We argue that the lack of relationship can be explained by the fact that what constitutes a "good" or "excellent" vocabulary is likely difficult to define since in-class feedback on vocabulary is not typically offered. While one learner may believe that an excellent vocabulary is one that allows for fluent daily speech, another may define it as sufficient for academic writing, or instruction of the target language.

Although an inability on the part of L2 learners to assess their proficiency has now been well-established, a shortage of research has looked into the many factors that might explain this phenomenon. We hypothesized that the personality trait, extraversion, would contribute to explaining variability in L2 self-assessed vocabulary since extraverts are more confident, self-assured, and optimistic than their self-critical, introverted counterparts. Indeed, we found support for this hypothesis via a significant positive correlation between extraversion and L2 self-assessed vocabulary ($r = .39$, $p = .00$). The hypothesis was further confirmed in a sequential regression model where extraversion explained 13% of the variance in self-assessed L2 vocabulary after accounting for objectively measured vocabulary. The fact that in the current study personality played a more important role in determining L2 self-assessments than actual ability, highlights the need to reconsider this tool for research, placement, and assessment purposes.

6. Limitations and suggestions for future research

The findings from the current study should be considered bearing in mind several important limitations. First, the measure of self-assessed L2 vocabulary was global, employing just one item. It is possible that a more specific, criterion-referenced measure of self-assessed L2 vocabulary would have lead to a relationship with objectively measured vocabulary that approached significance. We saw that between Brantmeier (2006) and Brantmeier et al. (2012), the relationship between self- and other-assessed reading strengthened and became significant with more specific self-assessment questionnaires. That said, even with extremely detailed self-assessment measures, correlations between self- and other-assessed skills remained weak. We therefore expect that if we had designed a more in-depth self-assessment instrument for vocabulary, self-assessments may have been more accurate, however not exact enough to consider the measure valid. Second, the self-assessment instrument did not distinguish between breadth and depth of vocabulary or receptive and productive vocabulary in asking learners to self-assess their L2 vocabularies. We chose vocabulary size as the objective measure upon the assumption that learners would be more inclined to focus on breadth than depth. Specifying between the parameters of breadth/depth and receptive/productive on the self-assessment instrument may have led to a stronger relationship between self- and other-assessed vocabulary.

Findings motivate further study of other variables, including individual differences, that lead to inaccurate perceptions of L2 ability. Multivariate research which measures a number of different independent variables within the same sample would be especially useful in order to provide insight into the relative contributions of different factors. This would bring us closer to a full

understanding of the variables that work together to explain differences in self-assessments.

7. Conclusion

While personality was a significant predictor of self-assessed L2 vocabulary, the intermediate-advanced learners of L2 French who participated in the current study did not accurately self-assess L2 vocabulary. Results not only emphasize the importance of personal variables in explaining self-assessments, but suggest that this tool may not be an appropriate way to evaluate L2 vocabulary. While criterion-referenced assessment tools may still be useful as a formative technique in classrooms, self-assessments should be avoided for high-stakes purposes. More generally, the absence of L2 self-perception observed here is problematic as it may impede learners' behaviour in L2 situations. We previously mentioned that self-assessments have been shown to influence confidence and willingness to communication in L2 classrooms (de Saint Léger, 2009; de Saint Léger & Storch, 2009). For example, over-confident learners may not participate in opportunities to improve lexical knowledge and may enter communicative situations which are too advanced (e.g., applying to be a French tutor). In contrast, those who are overly-critical may avoid L2 use due to self-consciousness and focus too much on developing vocabulary versus other L2 skills. In addition to these behavioural consequences of inaccurate L2 self-perception, Trofimovich et al. (2016, p. 134) have stressed that flawed self-assessment could be detrimental for L2 acquisitional processes according to interactionist approaches (e.g., Schmidt, 2001) which state that learners must notice gaps between their own and target language in order for L2 development to take place.

References

Alderson, Charles J. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.

AlFallay, Ibrahim. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, *32*(3), 407-425.

Anderson, Pamela L. (1982). Self-esteem in the foreign language: A preliminary investigation. *Foreign Language Annals*, *15*(2), 109-114.

Blanche, Patrick, & Merino, Barbara J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, *39*(3), 313-338.

Brantmeier, Cindy. (2006). Advanced L2 learners and reading placement: Self-assessment, computer-based testing, and subsequent performance. *System*, *34*(1), 15-35.

Brantmeier, Cindy, & Vanderplank, Robert. (2008). Descriptive and criterion-referenced self assessment with L2 readers. *System*, *36*(3), 456-477.

Brantmeier, Cindy, Vanderplank, Robert, & Strube, Michael. (2012). What about me?: Individual self-assessment by skill and level of language instruction. *System*, *40*(1), 144-160.

Burson, Katherine A, Larrick, Richard P, & Klayman, Joshua. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90,* 60–77.

Butler, Yuko Goto, & Lee, Jiyoon. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal*, *90*(4), 506-518.

Carter, Travis J, & Dunning, David. (2008). Faulty self-assessment: Why evaluating one's own competence is an intrinsically difficult task. *Social and Personality Psychology Compass*, *2*, 346-360.

Chamorro-Premuzic, Tomas, & Furnham, Adrian. (2005). *Personality and intellectual competence*. Mahwah, NJ: Lawrence Erlbaum.

Costa, Paul T, & McCrae, Robert R. (1997). Longitudinal stability of adult personality. In Robert Hogan, John A. Johnson, & Steven Briggs (Eds.), *Handbook of personality psychology* (pp. 269-290). San Diego, CA: Academic Press.

De Saint-Léger, Diane, (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals*, *42*, 158-178.

De Saint-Léger, Diane, & Storch, Neomy (2009). Learners' perceptions and attitudes: Implications for willingness to communicate in an L2 classroom. *System*, *37*(2), 269-285.

DeYoung, Colin G, Quilty, Lena C, & Peterson, Jordan B. (2007). Between facets and domains: Ten aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*, 880-896.

Dewaele, Jean-Marc. (2002). Psychological and sociodemographic correlates of communicative anxiety in L2 and L2 production. *The International Journal of Bilingualism*, *6*(1), 23-38.

Dlaska, Andrea, & Krekeler, Christian (2008). Self-assessment of pronunciation. *System*, *36*, 506-516.

Dörnyei, Zoltan. (2001). *Teaching and researching motivation*. London: Pearson Education Ltd.

Dörnyei, Zoltan. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Dunning, David, Meyerowitz, Judith A, & Holzberg, Amy D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082-1090.

Dunning, David, Heath, Chip, & Suls, Jerry M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5,* 69-106.

Foote, Jennifer. (2010). Second language learners' perceptions of their own recorded speech. *PMC Working Paper Series W P10-02*, 3-27. Edmonton: Prairie Metropolis Centre.

Funder, David C. (2001). Personality. *Annual Review of Psychology, 52,* 197-221.

Glover, Philip. (2011). Using CEFR level descriptors to raise university students' awareness of their speaking skills. *Language Awareness*, *20*, 121-133.

Hayes, Andrew F, & Dunning, David. (1997). Construal pro- cesses and trait ambiguity: Implications for self-peer agreement in personality judgment. *Journal of Personality and Social Psychology*, *72*, 664-677.

Hirsh, Jacob B, & Peterson, Jordan B. (2008). Predicting creativity and academic success with a "Fake-Proof" measure of the Big Five. *Journal of Research in Personality*, *42*, 1323-1333.

Juslin, Peter. (1994). The overconfidence phenomenon as a consequence of formal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, *57*, 226-246.

Little, David. (2005). The Common European Framework and the European Language Portfolio: Involving learners and their judgements in the assessment process. *Language Testing, 22,* 321-336.

Luoma, Sari. (2012). Self-assessment. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5169-5174). New York: Wiley-Blackwell.

MacIntyre, Peter D, Noels, Kimberly A, & Clément, Richard (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47*, 256-287.

McCrae, Robert R, & Costa, Paul T. (2003). *Personality in adulthood: A five-factor theory perspective* (2nd ed.). New York: Guilford Press.

McPeek, Robert, Nichols, Austin Lee, Classen, Sherrilene, & Breiner, Judith. (2011). Bias in older adults' driving self-assessments: The role of personality. *Transportation Research*, *14*, 579-590.

Ozer, Daniel J, & Benet-Martínez, Véronica. (2006) Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401-21.

Pervin, Lawrence A, & Cervone, Daniel. (2010). *Personality: Theory and research*. Hoboken, NJ: John Wiley.

Préfontaine, Yvonne. (2013). Perceptions of French fluency in second language speech production. *Canadian Modern Language Review*, *69*(3), 324-348.

Raasch, Albert. (1980). Self-evaluation in adult education. *Recherches et Echanges*, *5*, 85-99.

Ross, Steven. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 15,* 1-20.

Schmidt, Richard. (2001). Attention. In Peter Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Stankov, Lazar. (1998). Calibration curves, scatterplots, and the distinction between general knowledge and perceptual tests. *Learning and Individual Differences*, *8*, 28-51.

Suzuki, Yuichi. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing*, *32*(1), 63-81.

Thoresen, Carl J, Kaplan, Seth A, Barsky, Adam P, Warren, Christopher R, & de Chermont, Kelly. (2003). The affective underpinnings of job perceptions and attitudes: A meta-analytic review and integration. *Psychological Bulletin*, *129*(6), 914-945.

Trofimovich, Pavel, Isaacs, Talia, Kennedy, Sara, Saito, Kazuya, & Crowther, Dustin (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, *19*(1), 122-140.

Yaghoubi-Notash, Massoud. (2012). Holistic vs category-based self-assessment of expository L2 writing: Validity and reliability considerations. *International Journal of Applied Linguistics and English Literature*, *1*(2), 26-34.

Proceedings of the 42nd annual Boston University Conference on Language Development

edited by Anne B. Bertolini
and Maxwell J. Kaplan

Cascadilla Press   Somerville, MA   2018