Journal of Second and Multiple Language Acquisition – JSMULA
Vol: 8 Issue: 4 111-128, 2020, December
ISSN:4147-9747

Many-Facet Partial Credit Rasch Model Analysis of Dörnyei's L2 Motivational Self System

Received : 18.08.2020
Accepted : 27.10.2020
Published : 30.12.2020

Peter Kim[1]
Teachers College, Columbia University

Abstract

Motivation to learn a second language has been a significant predictor of variable success in L2 ultimate attainment. One prominent L2 motivation theory in SLA is Dörnyei's L2 Motivational Self System (L2MSS) in which learners' motivation is comprised of Ideal L2 Self, Ought-to L2 Self, and L2 Learning Experience, validated through their correlation with one's Intended Effort towards learning the target language. This paper aims to address two gaps in the L2 motivation literature. First is methodological: measures of motivation are often collected through questionnaires designed to probe the construct in question and analyzed using descriptive statistics or factor analysis. However, descriptive and inferential statistics often rely on interval data when motivation measures are likely to be ordinal. In this regard, the use of the Rasch model as a measure of motivation affords researchers a powerful means of comparing the motivation levels of participants as objective measures. Second, L2 motivation has been theorized as being dynamic and subject to changes throughout one's L2 learning experience (Dörnyei, 2010). This study examines a motivation questionnaire based on L2MSS and uses a partial credit model to analyze the interaction between length of study and learners' motivation as defined by the components of L2MSS. Ninety-one ESL students were given a 24-item motivation questionnaire. The result of the analysis showed that the most important aspect of L2MSS was the learners' learning experience. The best method of analysis for the Rasch model was to subdivide the questionnaire according to each of Dörnyei's subcomponents (i.e., Ideal L2 Self, Ought-to L2 Self, L2 Learning Experience, and Intended Effort) instead of analyzing the data as a single dimension of L2MSS. Finally, the findings showed no significant interaction between the length of study and learners' motivation. The analysis of overfit and underfit items and persons provides the advantage of using the Rasch model not found in structural equation models, factor analysis, inferential or descriptive statistics.

Keywords: Rasch Analysis, L2 Motivation, L2MSS, Survey Analysis, Many-Facet Partial Credit Model

1. Introduction

A common observable phenomenon in SLA is that second language learning in adults is variable in its outcome (Skehan, 1989). Variability in outcome among learners is attributed to differences in learner-internal characteristics, and among all individual difference (ID) variables, motivation is considered as one of the most important variables on par with aptitude in predicting ultimate attainment (Dornyei, 2005). In ID research, motivation's

[1] Peter Kim is a doctoral candidate in Applied Linguistics at Teachers College, Columbia University, specializing in second language acquisition. His research interests include second language aptitude and motivation. Correspondence: pk2505@tc.columbia.edu

importance is related to its utility as a strong predictor of second language acquisition success because motivation is often associated with intended effort that learners expend at various stages in their L2 development (Dörnyei et al., 2016; Gardner, 2000). For instance, Ehrman and Oxford (1995) examined the role of ID factors such as motivation, anxiety, language learning strategies, learning styles, self- esteem, and personality traits. The study identified aptitude as the factor that most strongly correlated with proficiency, followed by affective and motivation factors as the next highest correlates.

There is a priority of importance on motivation that is uniquely different from other individual differences such as personality, intelligence, or even aptitude. That is, a high level of motivation is seemingly able to overcome low aptitude or other disadvantages in individual differences. Without motivation, even the most talented learner cannot accomplish L2 mastery and neither quality teaching, curricula, nor the environment can make up for lack of motivation (Dörnyei & Skehan, 2003). As an aside, Dörnyei (2005) cites Robert Stenberg, a prominent psychometrician, whose comment illustrates the importance of language motivation in L2 learning. Stenberg noted that in Belgium those who learn Flemish are more likely to learn a 2nd and a 3rd language than those who learn French. Based on this trend, he argued that this was obviously not a phenomenon that should be attributed to aptitude but to motivation driven by need. Similarly, the often-ridiculed monolingual aspect of American society is not a reflection of American citizens' language aptitude — no one would assume that Americans in general have lower aptitude compared to Belgians — it is rather a reflection of motivation based on need regarding foreign language learning in America. Consequently, motivation is extremely critical for language acquisition, for without it how can language learning even begin? In fact, L2 learning presupposes motivation from the start and much of ID research is predicated on the assumption that the learner is motivated to pursue a second language (Ushioda, 2016). Much of research in SLA motivation has revolved around the question concerning the different types of motivation and how they are influenced by various factors that are both learner internal and external (for example see Dörnyei, 1996; Gardner & Lambert, 1965; Moskovsky et al., 2016; Williams & Burden, 1997).

1.1. Review of the Literature

Motivation to learn a second language has been widely acknowledged by researchers and language instructors alike as a key factor in the overall success of second language acquisition (Gardner, 2000; Skehan, 1989; Skehan, 1991). Not only does motivation provide the impetus to initiate L2 learning, it also provides the drive to sustain the pursuit (Dörnyei et al., 2016). R.C. Gardner (1985) defines motivation as a summation of intended effort with desire to achieve the L2 goal and the attitude towards the target language. Motivation in SLA research is described as a composite construct with components that appear to be individual attributes while others are situation-specific (Dewaele, 2009). The works of Gardner and Lambert during the 1950s and 1960s, also known as the social-psychology period of motivation, laid the foundation for the beginning of a large body of research

in L2 motivation. Their work on L2 motivation started out with a broad scope of social-psychological view by arguing that learning the language of another community cannot be separated from the learners' social dispositions towards the speech community in question. Gardner's social-educational model claims that what drives L2 learning is integrative motivation, which is the learner's willingness or desire to be like a representative member of the target language community. It is reflective of the extent to which the learner desires to be a part of the valued community and communicate with the target language group (Gardner & Lambert, 1965). Thus, motivation has societal influences and factors that are larger than the individual learner.

Following Gardner and Lambert's social-educational theory on motivation, a new generation of cognitive psychologists during the 1990s ushered in what Dörnyei and Ushioda call the 'cognitive-situated period' that focused on the individual (Dörnyei & Ushioda, 2009). As a result, L2 motivation theories have been increasingly interested in the socio-dynamic perspective that is concerned with ideations about the "self". Based on Markus and Nurius' (1986) possible-selves theory, and Higgins' (1987) Self-discrepancy theory, Dörnyei proposed L2 Motivational Self System for understanding L2 learner's motivation. Markus and Nurius's possible-selves theory stated that there are three possible-selves: what we would like to become, what we could become, and what we are afraid of becoming. Self-discrepancy theory as it applies to SLA argued that the discrepancy between what one would like to become and one's current state acts as a motivational springboard for learning a second language. Encompassing these theories is Dörnyei's L2 Motivational Self System (L2MSS). While Markus and Nurius' (1986) theory was not specific for language acquisition, Dörnyei's L2MSS addressed motivation as it pertains to learning a second language. The fundamental assumption in L2MSS is that L2 learners are driven towards the target language when they perceive a discrepancy between their ideal/ought-to self (future L2 using-self) and their current state (Dörnyei, 2005). There are three core components in Dörnyei's theory. First is the Ideal L2 Self, a desirable future self-image of the L2 user the person would like to become. The second is the Ought-to L2 Self which is the expectations imposed by others (e.g., family or society) that bears little resemblance to one's own ideal self. And the third is the L2 Learning Experience, which places the motivational impact of the learner's present learning situation, such as the instructor, peers, and the curriculum. According to this system, the three components reflect the L2 learner's perception of their identity and this perception operates in tandem with the influence of society and the environment to trigger motivation for L2 learning (Dörnyei & Ushioda, 2009). Measures of one's Ideal-Self, Ought-to-Self, and Learning Experience operate in concert to explain the variability found in their Intended Effort towards learning a second language. That is, motivation is able to explain or indicate the amount of effort put out by the motivated learner. Recently, a large-scale predictive validation study (N=360), examined the link between L2MSS and L2 achievement in reading and writing proficiency test. Analysis of results on L2MSS questionnaires revealed that the three main components of Ideal L2 Self, Ought-to L2 Self, and the L2 Learning Experience unequivocally had power to predict intended learning behavior. A meta-analysis of L2MSS was also recently reported by

Al-Hoorie (2018) based on 39 samples that totaled 32,078 language learners. In this study, the three components of L2MSS were found to be significant predictors of intended effort (rs=.61).

Invoking the dynamic systems theory (DST), Dörnyei argues that individual learner differences are dynamic, and they are comprised of different components that interact with each other and the environment. Dörnyei (2010) proposes that the learner differences be viewed as a complex dynamic system with interconnected components that fluctuate over time. Recent research in L2 motivation has begun to reflect this important aspect that motivation is process-oriented, never stable but subject to fluctuations as a state (Dörnyei, 2009). For example, a study of Iranian learners of English on demotivation revealed that the major source of demotivation during their EFL course was their inability to keep up with the class requirements and the rate of curriculum progression (Yaghoubinejad et al, 2016). Similarly, a study by Kruk (2016) showed that L2 motivation changed not only from class to class but also from one lesson to the next within the same class for reasons that included the nature of the activity, topic of the lesson, or even the use of the course book. A longitudinal study of first year German L2 students in UK revealed that although the students' desire to become proficient in German increased, their effort to engage the language decreased over the course of the year. Their decrease in effort coincided with decrease in the level of intrinsic motivation and self-efficacy beliefs (Busse & Walter, 2013). Summary of findings from studies of demotivation by Kikuchi (2015) identified experience of difficulty, followed by loss of interest and class environment as the best predictors of demotivation in adult English learners.

The most common standard of psychometric measure of motivation is questionnaire/survey. Motivation surveys come in different lengths, designed with specific purposes, for specific audience, use, and theoretical framework. They have been translated and used for a comparative survey in Japan, China, Hungry and Iran, as a scheme for classroom observation, to learn about motivational strategies, self-regulating capacity in vocabulary learning, language orientation survey, and to assess directed motivational currents in others (Dörnyei & Csizér, 2012). It is a common practice in social science research to evaluate the suitability of questionnaires based on reliability and validity using various statistical methods such as correlation analysis or factor analysis. L2 motivation measures have been assessed with similar standards in SLA research. Gardner and MacIntyre (1993) reported evidence of reliability and validity based on several multi-trait/multimethod analysis. Dörnyei's L2MSS questionnaires have also been vetted for the validity of its components through confirmatory factor analysis (Safdari, 2017) and structural equation modeling (Papi, 2010).

Although factor analysis and SEM have been implemented in the analysis of L2 motivation surveys, use of the Multi-Facet Rasch Model (MFRM) has been lacking in the literature. Rasch model is based on the premise that an individuals' responses to a class of items are based on the difficulty of the item and the said individual's ability, as measured by the item. When applied to measuring a single trait, the model assumes that people with high ability are more likely to get the items correct, and items that are the easier are more likely to be answered correctly than others (Boone & Noltemeyer,

2017). Rasch analysis uses logit scales (log odds) which can convert nonlinear raw scores to be linear, interval scale. The same concept can be extended for motivation survey data in which the individual's level of motivation comparable to others is determined by that individual's motivation and the probability of endorsing the questionnaire items that purports to measure the motivation in question (Bond & Fox, 2015). In performing linear model analysis with survey data, it is wrong to assume that Likert scales have equal intervals between them, so that the psychological distance between, for example, Strongly Agree to Agree is equal to the distance between Neutral to Disagree when such uniformity may not exist. MFRM makes possible the analysis of data that is ordinal such as survey data and convert them to be on equal-interval scales. This allows MFRM to produce measurable information about how survey items are performing with respect to the participants and creates a single frame of reference that facilitates comparisons within and between the variables or facets of the study (Boone, 2016).

In comparison to other methods such as SEM and factor analysis, Rasch model provides a distinct advantage. Structural equation models such as factor analysis (FA) are used to find the dimensionality of motivation questionnaire/surveys by identifying the constructs that give rise to the response patterns in the dataset. Likewise, Rasch analysis can be used to evaluate the dimensionality of items in a survey. However, unlike, factor analysis, which is based on a correlational model, Rasch analysis employs a hierarchical implicational model (Wright, 1996). The advantage of Rasch model is that more difficult to endorse items in the questionnaire are expected to be endorsed by learners who possess a greater level of motivation (Bond & Fox, 2015). In SEM/FA methods, 'endorse-ability' of Likert ratings in a given item can produce misleading results because easy to endorse items and difficult to endorse items will load onto different factors even if those items are designed to measure the same trait or construct of motivation. The critical advantage of Rasch analysis then, is that item difficulty is taken into consideration and their responses have probabilistic uncertainty. Items or persons that perfectly predict responses overfit the Rasch Model and indicate inefficient measure that provide no unique information, while items or persons that have high probabilistic uncertainty underfit the Rasch Model. These features allow for greater analysis and insight into the workings of items and responses which in turn provide deeper understanding of the motivational constructs explored by the questionnaire (Sick, 2011).

The focus of this study is on Dörnyei's L2MSS and the purpose is to address two gaps in the L2 motivation literature reviewed thus far. The first gap is methodological, as MFRM that evaluate ordinal data such as surveys has been lacking in SLA literature. Despite the advantages of applying Rasch measurement analysis in SLA research and their psychometric instruments, there is a significant lack of its use in SLA. Only about 2.01% of publications in linguistics have used Rasch measurement, making the field one of the bottom five adopters of Rasch analysis in social science (Aryadoust, et al., 2019). Given the lack of research in the area, it is unclear what would be the best approach to analyze a motivation survey made up of multiple

subsections. The second is theoretically motivated: investigating dynamic complexity of motivation and the length of L2 study in L2 learners. To that end, the following research questions are explored in the current study.

RQ 1. What do the components of Dörnyei's L2MSS reveal about the nature of L2 learner's motivation? (i.e., are learners mostly motivated by their Ideal-Self? Ought-to-Self?)
RQ 2. What is the best methodological MFRM approach to analyzing a survey with subsections?
RQ 3. Is there an interaction between length of study and learners' motivation as defined by the components of L2MSS?

2. Methodology

2.1. Participants

A total of 91 ESL students ranging in proficiency from beginner to advanced participated in the study. Participants were from an institutional ESL program in New York City. There were 22 males (27.5 percent) and 66 females (72.5 percent). The average length of years spent on studying English was 7.6.

2.2. Instrument

A 24-item questionnaire was designed to measure motivation according to the theoretical framework of Dörnyei's L2 Motivational Self System (L2MSS). Dörnyei's model was chosen because it has been one of most prominent theory that has dominated L2 motivation since its introduction in 2005 (Boo et al., 2015). The questionnaire comprised of items that measured Ideal L2 Self, Ought-to L2 Self, L2 Learning Experience, and Intended Effort. The Ideal L2 Self was measured by the extent to how the respondents envision themselves as English language users in their future. The Ought-to L2 self was measured by one's sense of obligations and duties the respondents felt in order to avoid the negative consequences of not learning the L2. The L2 learning experience was measured by considering their L2 learning context, environment, materials and instructors. Intended Effort measured the participants' willingness to go above and beyond the minimal requirement and extend effort in order to learn the L2. The items were adapted based on the questionnaire by Papi's (2010) structural equation modeling study and Islam and Chambers's (2013) survey of L2MSS. Each item required the respondent to rate how the statement accurately reflected their own opinions based on a 5-point Likert scale that ranged from strongly disagree, disagree, neutral, agree, and strongly agree. All items were positively worded in order to prevent reserve coding and also to align the numeric increase of Likert scale with increasing level of motivation indicated by the participants. This meant that responses of 5s indicated the highest level of motivation while responses of 1s signaled the lowest possible level of motivation towards learning English for all items.

2.3. Data collection

Data collection spanned three semesters and all participants were asked to volunteer in a research study. All participants took a brief background

survey and a motivation questionnaire. As part of the demographic survey, the number of years spent on studying English was also reported.

2.4. Data analysis

The data from the completed questionnaire were entered into an Excel spreadsheet, exported to FACETS (Linacre, 2007), and calibrated using the Rasch partial credit model. The 5-point Likert scale responses were entered as values 1 through 5. The length of time spent on studying English was dummy coded with a value of "1" for less than a year, "2" for between 1 to 5 years, and "3" for any length greater than 5 years. This resulted in 26 participants in the category of 1 (28.6 percent); 24 participants in the category of 2 (26.4 percent); and 41 participants in the category of 3 (45.1 percent). Three different MFRM analyses were performed on the survey data in order to triangulate the analysis of item, person, subsection, and length of study, and to compare the effectiveness of each method. The first model was a four-facet partial credit model for all items in the survey grouped by four subsections. The model is defined mathematically by the following formula:

$$\log\left(\frac{P_{nijk}}{P_{nij(k-1)}}\right) = B_n - D_i - C_j - F_{ik} \quad (1)$$

P_{nijk} is the probability that participant *n* endorses a Likert rating of *k* on item *i* by subsection *j*. $P_{nij(k-1)}$ is the probability that participant *n* endorses a Likert rating of *k-1* on item *i* by subsection *j*. B_n is the tendency of the participant *n* to report being motivated. D_i is the difficulty of endorsing item *i*. C_j is the difficulty of endorsing subsection *j*. F_{ik} is the difficulty (F) of endorsing a Likert rating *k* averaged across all items but for each subsection *j* separately.
The second model was a three-facet model without the grouping of subsections for all items in the survey, plus the length of study modeled as a dummy facet for the bias/interaction analysis. The model for the second analysis is defined mathematically by the following formula:

$$\log\left(\frac{P_{nik}}{P_{ni(k-1)}}\right) = B_n - D_i - F_k \quad (2)$$

P_{nik} is the probability that participant *n* endorses a Likert rating of *k* on item *i* $P_{ni(k-1)}$ is the probability that participant *n* endorses a Likert rating of *k-1* on item *i*. B_n is the tendency of the participant *n* to report being motivated. D_i is the difficulty of endorsing item *i*. F_k is the difficulty (F) of endorsing a Likert rating *k* averaged across all items.
The third model was identical to the second model, except that the data were split by subsections. Four three-facet models for only the items in each subsection was performed with the length of study bias/interaction analysis. The model for the third model is defined by the same formula as above (2),

iterated four times for Ideal-Self, Ought-to-Self, L2 Experience, and Intended Effort.

3. Findings

The results that follow are presented with respect to the three separate MFRM models specified above: Model 1 (full model with no interaction), Model 2 (no subsections + interaction), Model 3 (by subsections only + interaction).

Table 1
Model 1 Summary

	Person	**Item**	**Ideal Self**	**Ought-to-Self**	**L2 Experience**	**Intended Effort**	**Subsections**
Logit Spread	-1.44 to 3.56	-0.98 to 0.79	-0.72 to 0.79	-0.98 to 0.06	-0.03 to 0.65	-0.35 to 0.46	-0.26 to 0.45
Strata & Reliability	4.13 & 0.89	4.52 & 0.91	4.89 & 0.92	4.87 & 0.92	2.10 & .64	2.59 & .74	6.76 & .96
X^2	(90, N = 91) =771.0, p = .001	(23, N = 24) =297.1, p = .001	(5, N = 6) =71.8, p = .001	(5, N = 6) =60.3, p = .001	(4, N = 5) =11.1, p = . 03	(6, N = 7) =25.6, p = .001	(3, N = 4) =126.6, p = .0 01
Misfits	19, 48, 74, 13, 17, 18, 28, 41, 55, 14, 56, 39, 16, 10	7, 16		7			Ought-to-Self

The logit scale provides an equal interval representation for the participants on motivation measures so that they can be interpreted with respect to one another. Table 1 shows the logit spread of the person measures from -1.44 to 3.56 for a total spread of 5.00. A high measure indicates a correspondingly high level of motivation and a low measure indicates a lower level of motivation based on the 24-item questionnaire considered as whole. Only four participants had negative logits while the rest had positive logits. The majority had logits greater than 1 (55% of participants), indicating that, in general, the participants feel motivated to study English. This is hardly surprising for adult L2 learners who choose to take ESL classes by their own volition, unlike secondary school students who are often obligated to satisfy a foreign language requirement.

Figure 1 below also corroborates with the findings from Table 1 above. As expected, the most common Likert rating selected by the participants was 5 (strongly agree) followed by 4 (agree). "Strongly agree" comprised of 46.6 percent of the total responses, "agree" was 29 percent, "neutral" was 13.5 percent, "disagree" was 5.2 percent and "strongly disagree" was 5.7 percent

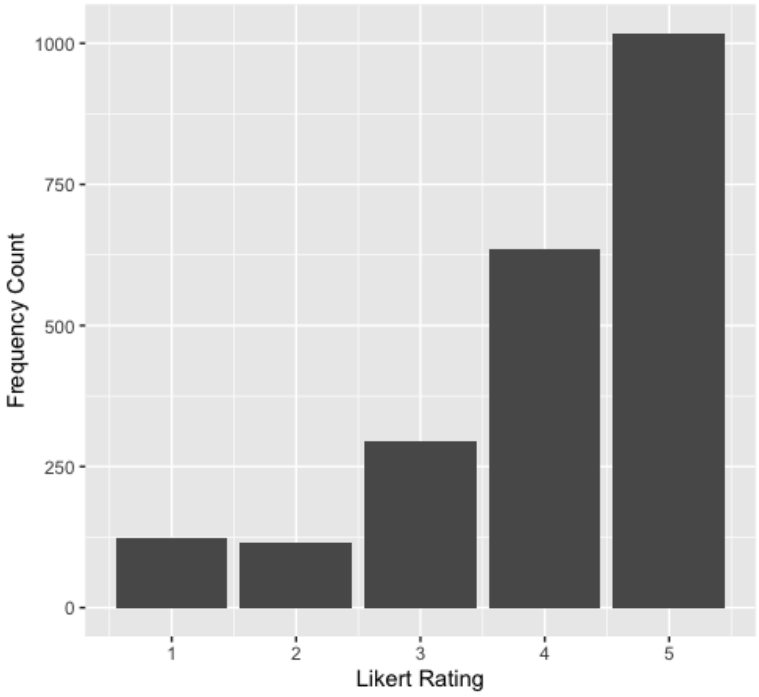

Figure 1. Bar plot of Likert Rating across 24 items

In Table 1, the strata statistic, which is the number of statistically distinct levels of participants' motivation in the sample of ESL students, was 4.13. This indicated that the survey was able to statistically distinguish about 4.13 levels of motivation in the participant group. Since the questionnaire used a 5-point Likert scale, 4.13-groups corroborates with the range provided by the scale. For the items, Table 1 shows the logit spread of the measures which ranged from -0.98 to 0.79 for a total spread of 1.77. Ten out of twenty-four items were found to have negative logits. Of those ten, five items belong to the subsection that measured Ought-to-Self. This suggests that the most difficult items to endorse (i.e., marks strongly agree or agree on the scale) were the items that pertain to learning English out of obligation or due to others' expectations. Taken together with the results from person measures, it shows that while most learners are motivated to learn English (i.e., have a positive logit), their source of motivation was not based on obligation or expectations of others.

Subsection 1, Ideal-Self, logit spread of the item measures ranged from -0.72 to 0.79 for a total spread of 1.51. Only two items had a negative logit and the rest were positive. The two items were: "I can imagine myself speaking English like a native speaker" and "I can imagine myself studying in a university where all my classes are taught in English." The negative logits on these items were reasonable because most participants have been studying English for years and they may feel that speaking like a native is not one of the most realistic outcomes. Also, not all students intend on going to college in U.S.; consequently, one may be motivated to learn but choose to not endorse either or both of these two statements.

Subsection 2, Ought-to-Self, logit spread of the item measures ranged from -0.98 to 0.06 for a total spread of 1.04. Five out of six items had negative

logits. Clearly, this was the most difficult construct to endorse by the participants (Appendix A). It seems that participants wanted to indicate that learning English was their choice and not based on a sense of duty towards others. Interestingly, the only item with the positive logit was the statement "studying English is important to me because other people will respect me more if I have knowledge of English." This was the only statement that did not explicitly mention expectations of other people but state how learning English may improve one's own sense of self-worth as viewed by others. If one's Ideal-Self and Ought-Self is seen to be on a continuum, this item may lie closer to the learner's sense of Ideal-Self than other items since it reflects how they want to be viewed, ideally. Item 7 "I study English because friends and family think it's important", was found to be misfitting with an Infit Zstd of 3.0. Item 7 was also one of the most difficult items to endorse among Ought-to-Self items and the underfit of this may be an indication that highly motivated individuals do not endorse the statement: "I study English because friends and family think it's important." This item was the only item that mentioned family and friends, and this could indicate that for adult learners who are motivated, external pressures by friends and family have very little effect on one's overall source of motivation, which may not be the case for younger learners.

In subsection 3, Learning Experience, the logit spread of the item measures ranged from -0.03 to 0.65 for a total spread of 0.68. Only item 16 "Time goes by fast when I am studying English" was negative (-0.03). The facet map in Appendix A shows that this component/subscale was the easiest to endorse by the participants. In general, most participants reported strongly positive L2 learning experience. However, this component is also most likely influenced by the learning environment such as the quality of the instructor, learning materials, and the learning institution. This is reflected in the lowest reliability measure among all constructs, at .64, which indicated some variable responses in the questionnaire for this subsection. In addition, the survey was taken during an ESL class while the instructor was present, and this raises doubts about whether the participants were being conscientious about their answers. Interestingly as previously stated, the lone negative logit statement was "Time goes by fast when I am studying English." This statement may have been misunderstood by low proficiency participants as a negative statement or that "times goes by fast" was understood to mean not enough time.

Subsection 4, Intended Effort, the logit spread of the item measures ranged from -0.35 to 0.46 for a total spread of 0.81. Two out of seven items had negative logits (items 18 and 20, at -0.05 and -0.35, respectively) meaning that this construct was overall rather easy to endorse, on par with Ideal-Self but not as much as L2 Learning Experience. Items 18 and 20 stated the following: "I would like to spend a lot of time studying English" and "I would like to study English more than any other topics." This could indicate that although the learners are willing to study English, they are not intrinsically motivated, meaning that they are not motivated to learn English simply for the sake of learning the language but for other instrumental reasons. The four subsections considered together as a facet had a logit spread that ranged from -0.26 to 0.45 for a total spread of 0.71. Only the Ought-to-Self

had a positive logit, meaning that out of the four components, Ought-to-Self was the construct most likely *not* to be endorsed by the participants as previously discussed. Ought-to-Self was also the only misfitting construct at Infit Zstd of 2.6, most likely due to the underfitting of Item 7 ("I study English because friends and family think it's important"), and also due to the fact that it was the only construct whose elements were mostly negative in logits.

Table 2
Model 2 Summary

	Person	**Item**
Logit Spread	-1.37 to 3.55	-1.09 to 1.37
Strata & Reliability	4.12 & 0.89	6.25 & 0.95
X^2	(90, N = 91) =718.3, p = .001	(23, N = 24) =577.1, p = .00 1
Misfits	74, 48, 13, 18, 19	30, 82, 16, 41, 17
Bias/Interaction	item 19 (effort) with <1 t= -2.17, p=0.04	

Table 2 shows the key summary statistics of Model 2, which was identical to the first model but without the facet of subsection. Interestingly, the number of person misfit decreased but the number of item misfit increased. When participant's motivation is considered based on all 24 items, meaningful variance from the model may have decreased since the item facets are no longer divided into subsections. However, when items are considered together as a single motivation construct, this resulted in more misfits. One possible explanation is that due to the contradicting functioning of Ought-to-Self relative to other constructs, there were more items that appeared to be misfitting. While the second model assumed that the 24 items were unidimensional, they were in fact comprised of four dimensions. Failure to account for the difference in dimensionality may have led to more items being misfitting. Model 2 showed one statistically significant interaction between Item 19 (effort) with length of study less than one year (t= -2.17, p=0.04). However, the t-statistics was barely over 2 and the fact that only one item was found to interact indicates that there is likely to be no substantial interaction between item and length of study.

Table 3
Ideal-Self only

	Person	**Item**
Logit Spread	4.82 to 5.30	-1.26 to 1.49
Strata & Reliability	2.52 & 0.73	6.72 & 0.96
X^2	(90, *N* = 91) =391.7, *p* = .001	(5, *N* = 6) =4.8, *p* = .001
Misfits	30, 82, 16, 41, 60, 17	
Interaction	item 2 with <1 t= -2.73, p=0.0147	

Table 4
Ought-to-Self only

	Person	**Item**
Logit Spread	-3.84 to 4.36	-0.82 to 0.81
Strata & Reliability	3.08 & 0.81	6.07 & 0.95
X^2	(90, *N* = 91) =387.9, *p* = .001	(5, *N* = 6) =93.5, *p* = .001
Misfits	31, 74, 12, 51, 48, 40, 56,	12
Interaction	None	

Table 5
L2 Experience only

	Person	**Item**
Logit Spread	-2.67 to 6.16	-0.57 to 0.58
Strata & Reliability	2.58 & 0.74	2.88 & 0.79
X^2	(90, *N* = 91) =357.6, *p* = .001	(4, *N* = 5) =18.6, *p* = .001
Misfits	36, 39, 54, 88, 90, 20, 8	None
Interaction	None	

Table 6
Intended Effort only

	Person	**Item**
Logit Spread	-2.35 to 4.85	-0.46 to 0.66
Strata & Reliability	2.35 & 0.70	3.12 & 0.81
X^2	(90, *N* = 91) =324.8, *p* = .001	(6, *N* = 7) =35.3, *p* = .001
Misfits	83, 61, 39, 57, 88	None
Interaction	item 20 with <1 t= 2.16, p=0.0447	

Table 3 shows the summary statistics of Ideal-Self only data. Logit spread was from -4.82 to 5.30 for a total spread of 10.12. Twenty participants had maximum possible extreme score and one participant had the minimum possible extreme score; thus, their fit statistics could not be computed. There was one statistically significant interaction between Item 2 and a group of learners with less than a year of studying English. Item 2 states "I can imagine myself speaking English with international friends." Beginners were more likely to agree with this statement than more senior groups, which reveals the language use realities of the more versus less experienced learners. In other words, beginners are more optimistic about their ultimate attainment (i.e., being able to speak with ease in English) compared to advanced learners with more years of English education.

Table 4 shows Ought-to-Self only: logit spread was from -3.84 to 4.36 for a total spread of 8.2. Seven participants had maximum possible extreme score

and three participants had the minimum possible extreme score; their fit statistics could not be computed. Table 5 shows Experience Only with logit spread from -2.67 to 6.16 for a total spread of 8.83. Thirteen participants had maximum possible extreme score; their fit statistics could not be computed. Table 6 is Effort Only summary statistics. Logit spread was from -2.35 to 4.85 for a total spread of 7.2. Fourteen participants had maximum possible extreme score; their fit statistics could not be computed. There was one statistically significant bias/interaction of item 20 with a group of learners with less than a year of studying English. Item 20 states "I would like to study English more than any other topics." Beginners were more likely to agree with this statement than more senior groups. It is possible that beginners are more enthusiastic towards studying English than other topics because it is a relatively new pursuit for them.

4. Discussion

Even when anonymity is guaranteed and there is no incentive to answer in any particular way, responses to motivation questionnaire can be biased to make participants look studious or to confirm to the expectations of the instructor/researcher. L2 motivation questionnaires naturally inquire about the students' desire to learn the TL, how they feel about the target language community, and how much they are devoted to their language acquisition goal. Given that most respondents were adult learners who voluntarily chose to enroll, it is somewhat expected that they would answer favorably towards the target language, the learning environment, idealizations about the language culture, and the language-speaking community. This was reflected in 75.6 percent of the responses as being either "strongly agree" or "agree" to the positively worded items (Figure 1).

In light of this tendency, there are few reasons why person misfit on survey data are difficult to interpret. First, there is a general tendency for people to want to present a favorable image of themselves in questionnaires, known as socially desirable responding (SDR) (Mortel, 2008). Socially desirable responding and wanting to meet expectations of the researcher can cause respondents to report an inflated level of their own motivation. This means that the measure of motivation may be biased towards overestimation due to the assumed expectations and wanting to portray themselves as motivated learners. Moreover, since there is no "correct" answer for each item, each participant's measure of motivation is based on how they choose to answer and not based on whether their answer was an accurate measure of their true motivation level (which can never be known apart from a secondary criterion validation). Despite these challenges, extreme misfits do avail some interpretability. For example, both person number 19 and 48 had the highest underfitting of Infit Zstd of 3.6. Person 19's response to 24-items were mostly 1s and 3s except item 8 which had a response of 5. Closer examination shows that a response of 5 to item 8 "If I fail to learn English, I'll disappointment lots of other people" contradicts a neutral response of 3 on item 7, "I study English because friends and family think it's important." However, it is also plausible that this is not contradictory at all. The expression "other people" in item 8 do not necessarily include "friends or family" though that interpretation is somewhat unlikely. A more plausible

explanation is that even if the two terms are indeed equal, Person 19 might have felt sure that failing in English will disappoint his/her friends or family but still felt ambivalent about whether he/she was studying English because friends and family think it's important. A similar phenomenon was observed for Person 48. Person 48's responses were mostly 5s, expect items 9 ("Learning English is important because the people that I respect think that I should learn English"), 11 ("Lots of people expect me to learn English"), and 12 ("Studying English is important to me because other people will respect me more if I have knowledge of English"), which were all marked as 1s on Ought-to-Self statements. Person 48 strongly disagreed that he/she was learning English because of the expectations of others or because other people will respect him/her more. These responses appear to contradict 5s on an item that says "I study English because friends and family think it's important" and also on an item that says "Studying English is important in order to gain approval of my peers/teachers/family/boss." However, once again in the participant 48's mind, these statements may be referring to separate situations and not indicative of the same phenomenon. The case of participant 48 shows an example of important diagnostic analysis of individual participant afforded by the Rasch model because in most models, individual anomalies like these would be averaged away into the data, or just considered problematic as "outliers" and omitted from the dataset altogether.

To review, the first research question asked about the motivation makeup of ESL students based on Dörnyei's L2MSS. The sample of ESL students in this study demonstrated that their source of motivation was mostly from the Learning Experience and the quality of instruction received in the ESL classes. The Ought-to-Self was the least motivating source for the participants. Another research question was about the role of length of study on learners' motivation as defined by the components of L2MSS. It appears that there were not enough significant interactions between the items and the length of study to support the claim that motivation is dynamic. Thus, based on this set of data, motivational makeup of student who had more years of studying English did not appear to be different from those who were beginning learners.

One methodological implication of this study was on how the Rasch model can be used in survey data comprised of sections that measure separate constructs. The best method for analyzing survey data was to include the subsections as a separate facet. Model 1 provided the most useful information and insight into understanding the functioning of items that purported to measure different elements of motivation. Model 2 provided similar information regarding person and item fit as model 1, but an analysis of subsections was missing. In addition, item misfits on Model 2 were misrepresentative of the underlying composite constructs that make up the survey. Finally, Model 3 proved to be the least useful model in terms of interpretability. By dividing the total number of items into its composites and running the analysis as separate sample, a substantial amount of person measures could not be computed. A number of logit measures became maximum or minimum values because there were not enough items in each subsection to prevent maximum or minimum value responses. This meant key summary statistics of four separate iterations of Model 3 could

not be properly interpreted since maximum and minimum measures do not contribute to Rasch model estimates. It also reduces the data so much that it makes the construct within each component too restricted in a theoretical sense and does not allow for much room for measurement error. Finally, from a pedagogical standpoint, this study highlights the importance of the learner's L2 experience established through the teacher. According to the findings of this study, the role of the instructor and the quality of instruction were extremely important in influencing the motivation of the learner. Experience was the most influential source of motivation which was greater than the role played by one's family obligations, intended effort, and their imagined ideal self. This is a commonly supported phenomena in SLA literature; for example, Ushioda (2005) investigated the role of motivation for a self-paced online language course, and the results showed that the instructor was influential in affecting student's motivation to learn.

5. Conclusions

Overall, there were two limitations of the study. The first was the limited knowledge of English for low proficiency level participants. Because the survey was in English, for beginner level participants the meaning of the items may have been unclear and this could have results in inaccurate response in the data which could have led to either overfitting (e.g., series of 5s across the board) or underfitting (random responses). The second limitation was the implementation of length of study as a dummy variable. By dummy coding the participants' length of study into three levels, the continuous ratio scale of length in years was forced to become nominal. Unlike gender, race, or political affiliation, length in years is more appropriately considered as interval data, thus dummy coding the variable comes at the price of losing continuous interval information. A better way to examine the interaction effect of items with length of study would have been to keep the variable as a numeric ratio, which could have revealed more about its interaction with the functioning of the items.

The motivation survey examined in this study showed that learners are heavily influenced by their learning experience. In fact, the learning experience may be more important than one's Ideal-Self, or one's Ought-to-Self. The finding speaks to the importance of instruction and the quality of the learning environment as one of the most important motivating sources for learners. Regarding the dynamic nature of motivation, not enough evidence was found in this study to indicate that length of study interacted with motivation to any great nor systematic degree. This could be a direction for future longitudinal research since cross-sectional data are not accurate representation of developmental changes through time.

Lastly, given the literature on L2 motivation, surveys should be broken into subsections that are composites of the L2 motivation theory. Hence, MFRM by subcomponents provided the best analysis. A unidimensional measure of motivation is not as useful or insightful as a composite understanding of parts (or subsections) that work together to create learners' overall level of motivation. Division of survey data by section, if appropriate, decreases item misfit but may increase person misfit due to the conflicting dimensionality of the survey's underlying structure. Finally, interpretation of person misfit on

motivation surveys is subjective because there is no single "correct" answer and seemingly contradicting statements can have plausible interpretations. Overall, this study demonstrates MFRM as a powerful diagnostic tool for analyzing how the items in questionnaire/survey within the scale are working together according to the theoretical constructs it purports to measure.

References

Al-Hoorie, A. H. (2018). The L2 motivational self-system: A meta-analysis. *Studies in Second Language Learning and Teaching*, *8*, 721-754.

Aryadoust, V., Tan, H. A. H., & Ng, L. Y. (2019). A Scientometric review of Rasch measurement: The rise and progress of a specialty. *Frontiers in Psychology*, *10*, 2197. https://doi.org/10.3389/fpsyg.2019.02197

Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System*, *55*, 145–157.

Boone W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How?. *CBE life sciences education*, *15*(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1), 1416898.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in human sciences (3rd ed)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Busse, V. & Walter, C. (2013). Foreign Language Learning Motivation in Higher Education: A Longitudinal Study of Motivational Changes and Their Causes. *Modern Language Journal*, *97*, 435–456.

Dewaele, J. (2009). Individual differences in second language acquisition. In Bhatia, T. K., & Ritchie, W. C. (Eds.). *New handbook of second language acquisition* (pp. 623-646). New York, NY: Academic Press, Inc.

Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Dörnyei, Z., & Csizér, K. (2012). How to design and analyze surveys in SLA research? In A. Mackey & S. Gass (Eds.), Research methods in second language acquisition: A practical guide (pp. 74-94). Malden, MA: Wiley-Blackwell.

Dörnyei, Z., Henry, A., & Muir, C. (2016). *Motivational currents in language learning: Frameworks for focused interventions*. New York: Routledge.

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.

Dörnyei, Z. (2009). Individual Differences: Interplay of Learner Characteristics and Learning Environment. *Language Learning, 59*(1), 230–248.

Dörnyei, Z. (2010). The relationship between language aptitude and language learning motivation: Individual differences from a dynamic systems

perspective. In E. Macaro (Ed.), *Continuum companion to second language acquisition* (pp. 247-267). London: Continuum.

Dörnyei, Z., & Ushioda, E. (2009). *Motivation, Language Identity and the L2 Self*. Buffalo, NY: Multilingual Matters.

Ehrman, M.E., & Oxford, R. L. (1995). Cognition plus: correlates of language learning success. *The Modern Language Journal*, *79*, 67–89.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.

Gardner, R.C. (2000). Correlation, causation, motivation, and second language acquisition. *Canadian Psychology, 41,* 10-24.

Gardner, R. C., & Lambert, W.E. (1965). Language aptitude, intelligence, and second-language achievement. *Journal of Education Psychology 56*, 191-199.

Gardner, R. C., & MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, *43*(2), 157-194.

Higgins T. (1987). Self-Discrepancy: A Theory Relating Self and Affect. *Psychological Review*, *94*(3), 319–340.

Islam, M., Lamb, M., Chambers, G. (2013). The L2 Motivational Self System and National Interest: A Pakistani perspective. *System*, *41*(2), 231–244.

Kikuchi, K. (2015). *Demotivation in Second Language Acquisition. Insights from Japan*. Bristol: Multilingual Matters.

Kruk, M. (2016). Temporal fluctuations in foreign language motivation: Results of a longitudinal study, *Iranian Journal of Language Teaching Research, 4*(2), 1–17.

Linacre, J.M. (2007). *A user's guide to WINSTEPS*. Chicago: winsteps.com.

Markus, H., & Nurius, P. (1986). Possible Selves. *American Psychologist,* 41(9), 954–969.

Mortel, T. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, *25*(4), 40–48.

Safdari, S. (2017). Validation of a tailored L2 motivational self-system questionnaire through confirmatory factor analysis. *International Online Journal of Education and Teaching (IOJET)*, *4*(2). 174–183.

Skehan, P. (1989). *Individual differences in second-language learning*. London: Arnold.

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition, 13*, 275-298.

Papi, M. (2010). The L2 Motivational Self System, L2 Anxiety, and Motivated Behavior: A Structural Equation Modeling Approach. *System*, *38*(3), 467–479.

Ushioda, E. (2005). The Role of Students' Attitudes and Motivation in Second Language Learning in Online Language Courses. *CALICO Journal*, *23*(1), 49–78.

Ushioda, E. (2016). Language learning motivation through a small lens: A research agenda. *Language Teaching*, 49(4), 564-577.

Yaghoubinejad, H., Zarrinabadi, N. & Ketabi, S. (2016). Fluctuations in Foreign Language Motivation: An Investigation into Iranian Learners' Motivational Change over Time, *Current Psychology*,1–10.

Appendix A: Facet Map

Vertical = (1*,2*,3A,S) Yardstick (columns lines low high extreme)= 0,8,-2,4,End