

Research Article

EXPLORING THE RELATIONSHIP BETWEEN SECOND LANGUAGE LEARNING MOTIVATION AND PROFICIENCY A LATENT PROFILING APPROACH

Karen Dunn* 

Assessment Research Group, British Council, London, UK

Janina Iwaniec 

Department of Education, University of Bath, Bath, UK

Abstract

A foundation of second language motivational theory has been that motivation contributes to explaining variance in language learning proficiency; however, empirical findings have been mixed. This article presents an innovative approach to exploring L2 proficiency and motivations of teenage English language learners in Madrid, Spain (N = 1773). Participants completed a multiskill English language test, plus an eight-scale questionnaire operationalizing constructs from Dörnyei's L2 Motivational Self System (Dörnyei, 2005). Data were analysed using Latent Variable Mixture Modeling, a person-centered profiling approach. Results indicated five distinct classes of students, characterized by differing motivation-proficiency profiles. The importance of this study is that the analysis does not assume a homogenous relationship between motivational traits and proficiency levels across the learner sample; whilst there is undoubtedly a connection between the two areas, it is not a straightforward correlation, explaining to some extent discrepancies in previous findings and laying groundwork for further, more nuanced, investigation.

The authors would like to extend thanks to the British Council English Impact 2017 team, as well as colleagues at the British Council Madrid, without whom the extensive and valuable dataset used in the analysis would not exist. We would also like to recognize the very useful contributions from the anonymous peer reviewers for helping to shape the final article, as well as Professor Xiao Lan Curdt-Christiansen and Professor Barry O'Sullivan for reading early drafts and insightful comments along the way.

 The experiment in this article earned an Open Materials badge for transparent practices. The materials are available at <https://www.iris-database.org/iris/app/home/detail?id=york:939629>

* Corresponding author. E-mail: karen.dunn@britishcouncil.org

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

INTRODUCTION

A foundation of motivational theory is the supposition that motivation contributes to explaining variance in second language learning proficiency; in some social contexts it has been presented as the main driver of increased achievement (Gardner & Lambert, 1972). More recent research has been devoted to exploring constructs from the L2 Motivational Self System (L2MSS) (Dörnyei, 2005), with the most recent ones delving deep into the conceptualization of its key notions, ideal L2 self and ought to-L2 self (e.g., Feng & Papi, 2020; Papi et al., 2019; Papi & Khajavi, 2021). Whilst there is a body of research demonstrating the relationship between measures of effort invested in language learning and L2MSS constructs, caution needs to be taken in interpreting and comparing findings, as often the measures of proficiency used are not stringent, or rely on self-reported intended effort (cf. Al-Hoorie [2018]).

Several recent studies meanwhile provide sophisticated insights into the multiple and diverse motivational profiles of learners (Csizér & Dörnyei, 2005; Kangasvieri, 2017; Papi & Teimouri, 2014). Findings indicate a diversity in the levels of a range of motivational traits for different groups of language learners. However, in exploring the motivation-proficiency relationship, these studies tended to employ mainly correlational analyses, which imposes a simplified view of the relationships at play from the outset. We believe there is a clear need to explicitly integrate recent understandings of motivational profiles into explorations of how motivation relates to proficiency. To this end, the present study integrates a robust measure of language proficiency into a motivational profile analysis using a *person-centered* analytic approach. This escapes the *variable-centered* assumption that the nature of the relationship between motivational traits and proficiency will be the same for all learners in the sample. The methodology employed, Latent Variable Mixture Modeling (LVMM), is consistent with previous research into motivational profiles, and, by integrating a fully validated measure of English language proficiency, is open to nuances in the motivation-proficiency relationship that might otherwise have been overlooked.

LITERATURE REVIEW

LANGUAGE LEARNING MOTIVATION

Research on language learning motivation has a long history. In the current study, we adopt Dörnyei's L2MSS as the main framework through which to explore motivational influences, but also draw on other constructs to provide what we believe to be a more complete picture of the different aspects of motivation. The L2MSS proposes that the decision to pursue and sustain language learning might be underlined both by a vision of what the individual *wants* to achieve, coined as the "ideal L2 self," and individual's belief of what one *should* aspire to, referred to as "ought-to L2 self" (Dörnyei, 2005). Another component that affects these decisions is "language learning experience" that concerns the effect of the immediate environment on language learning. The key motivational force in Dörnyei's theory is driven by the gap between the future selves (ideal and/or ought-to) and current self-perception. The latter corresponds to English self-concept (Shavelson et al., 1976), which reflects how the learner views their L2 self at the present moment. Other constructs drawn on in this study reflect learning goals. The most commonly

researched language learning goals are “instrumentality,” which concerns the utilitarian reasons for learning a language (Gardner & Lambert, 1972) such as better job prospects, and “international orientation,” where English is learned to be able to communicate with other English speakers, and learn more about other cultures (Yashima, 2000).

STUDIES LINKING LANGUAGE LEARNING MOTIVATION AND LANGUAGE PROFICIENCY

Early studies of language learning motivation revealed a link between language learning motivation and proficiency, in particular motivation underlined by integrative motives, that is the willingness to be like a valued member of the target language community (Anisfield & Lambert, 1961; Gardner, 1960; Gardner & Lambert, 1959, 1972; Peal & Lambert, 1972). However, this body of research had substantial methodological limitations. Studies employed a mixture of measures, some of which were highly subjective or only tenuously related to language proficiency. Examples include: school grade reported by students; number of years speaking the language; average grade across a wide range of courses; and teachers’ rating of achievement (Lambert et al., 1972; Peal & Lambert, 1972). A study in the 1990s by Gardner et al. (1997) found motivation to be, in fact, more closely related to grades allocated by teachers than other more objective measures of achievement.

The line of research exploring the link between motivation and language proficiency appears to have been virtually abandoned, with few studies on the topic published after the initial wave in the 1960s and early 1970s. A number of studies since 2010 do, however, explore the relationship of L2MSS constructs with L2 proficiency (Dörnyei & Chan, 2013; T.-Y. Kim & Kim, 2014; Y.-K. Kim & Kim, 2011; Lamb, 2012; Moskovsky et al., 2016; Papi & Khajavy, 2021; Saito et al., 2019). Yet often the measures of proficiency are not robust, which makes comparisons difficult. Of these studies, in fact only a handful have the central aim to investigate the link between motivation and proficiency. These deserve some more attention here. Lamb’s study (2012) employed the C-test to measure Indonesian students’ proficiency level. Regression analysis revealed that only one motivational variable positively contributed to explaining variance in the test results, namely language learning experience at school. The role of the self-guides (ideal and ought-to) was, however, not clear. Moskovsky et al. (2016) used reading and writing tasks to measure the proficiency of English majors at a Saudi university. Correlation analysis showed the ideal L2 self to have a negative relationship with test results, with nonsignificant correlations for ought-to L2 self and language learning experience. The third study, which was longitudinal, focused on oral skills, and employed comprehensibility judgments by professional raters to measure proficiency (Saito et al., 2019). Similar to the two previous studies, Saito et al.’s (2019) results pointed to a negligible role of the ought-to L2 self in explaining oral proficiency gains of Japanese learners of English but the ideal L2 self was found to be positively related to proficiency. Papi and Khajavy (2021) explored this further and showed that ideal L2 self and ought-to L2 self with different regulatory focus (own/other) lead to qualitative differences in motivated behavior that in turn affect achievement in different ways. Clearly there are nuances in the motivation-proficiency relationship that are not being captured by existing approaches to researching this area.

Al-Hoorie's meta-analysis which incorporated 39 unique samples and 32,078 language learners (Al-Hoorie, 2018) pointed to some of the conceptual difficulties in comparing studies. Such discrepancies are understood to have led to an overall unclear picture of relationships between L2 language proficiency and the motivational traits specified by contemporary theorists. There are also differences in the research context and type of analysis employed, as well as a focus on different language skills, and use of different measures of proficiency. It is clear that there is a need for further comprehensive studies that clarify the relationship between constructs in the L2MSS and L2 achievement.

MOTIVATIONAL PROFILING

The profiling approach can add to our understanding of the relationship between L2 motivation and L2 proficiency. Profiling is a person-centered methodology, which assumes heterogeneity in the population studied, bringing advantages for the level of analytic insight for a given dataset (Marcoulides & Heck, 2013). Simply put, profiling is a means of grouping study participants based on their comparative levels across a series of metrics. For example, in terms of L2 skills, a profiling approach may show that one group of students performs comparatively better in productive over receptive skills, whilst the reverse may be true for a different group of students. In this case looking at learner profiles helps to qualitatively discriminate L2 ability for different groups of students. Such classification and understanding of the characteristics of students of different profiles opens up avenues for future interventions, changes in curriculum, or instructional modifications to provide bespoke support for different groups of learners.

Profiling has been used in SLA research to explore the interrelationships between motivational traits. Four key studies are relevant to mention here, three of which used cluster analysis to establish patterns of motivational trait levels: Csizér and Dörnyei (2005) studied motivation profiles of Hungarian eighth-grade language learners; Yun et al. (2018) worked with college students from South Korea; and Papi and Teimouri (2014) Iranian secondary school students. A further study by Kangasvieri (2017) employed latent profile analysis (LPA) to investigate motivational profiles of Finnish secondary school students learning a range of foreign languages. These four studies are not directly comparable to one another because they employed differing motivational variables and methodologies; there are, however, two clear parallels between their results. First, each study identified four or five distinct motivational profiles, with contrasting profiles of students with weak and strong motivation. Second, distinctions between groups were not always uniform across the traits, for example Kangasvieri (2017) revealed self-concept in particular to vary independently of other motivational traits between the different profile groups.

Of the four studies discussed here, only Yun et al. (2018) incorporated a measure of proficiency (Test of English Proficiency [TEPS], developed by Seoul National University), which enabled them to identify five distinct learner profiles: Thriver, Engaged, Striver, Dependent, and Disengaged. The possibilities engendered by integrating proficiency measures into motivation profiles is also emerging in other fields of educational assessment. A key example is from Michaelides et al. (2019) who use cluster analysis to explore the relationship between motivation and achievement in mathematics across a

range of contexts. These researchers emphasize that a person-centered analytic approach enables motivational predictors of achievement to “interact in ways that will reveal stronger and weaker associations with achievement, compared to the average relationships observed in variable-centred approaches” (Michaelides et al., 2019, p. 4). The study reported here is likewise premised on the possibility that the strength, or even the directionality, of motivation-proficiency relationships can vary between learners and learner groups. In taking a person-centered profiling approach in modeling L2 motivation and integrating robust measures of L2 proficiency, the study described in this article aims to set the groundwork for a more comprehensive and intricate understanding of the relationship between motivation and proficiency than has come before.

RATIONALE AND RESEARCH QUESTIONS

This study takes steps toward addressing acknowledged gaps in our understanding of the relationship between L2 language learning motivation and proficiency. The review of literature suggests that multiple motivational subpopulations exist within a given group of teenage English language learners, each with a distinct profile. Notably, it has also been shown that integrating proficiency information into this profiling approach has the potential to add a valuable dimension to the findings. The current research progresses our understanding by integrating information from a validated language test into a rigorously executed and methodologically robust motivational profiling study. The following research questions are addressed:

Are there heterogenous subgroups of students within the “English Impact” Madrid data? If so, how do these profiles differ with respect to L2 proficiency and motivational traits?

To answer these questions, the data collected for the British Council’s English Impact Madrid study (Shepherd & Ainsworth, 2017) was revisited to establish the number and nature of subpopulations in this learner group with respect to motivation and proficiency. The data comprised of 15.5-year-old English language learners’ answers to a motivational questionnaire and the results of a multiskill English language test; all learners were currently attending state schools in the Madrid region of Spain. It is hoped that these results will be a basis for further research that will ultimately lead to a better understanding of the needs of different groups of learners and how they can be addressed by instructional intervention and curricular reform.

METHOD

SAMPLING AND PARTICIPANTS

Data were collected as part of a wider study run by the British Council in 2017 (Shepherd and Ainsworth, 2017). Participants from within the Madrid state school system were selected using two-stage cluster sample methodology designed and executed by the Australian Council for Educational Research (ACER). First a sample of schools based on defined stratification variables (school type, geographical location, and bilingual

status) was chosen, and then a random selection of 12 students from each school was sampled.

Participating students were from Compulsory Secondary Education (ESO) 4 and their mean age was 15.6. They were currently studying English as part of their studies at this grade level for at least 90 minutes per week. In total, 1,773 students from 169 schools participated, 50.9% of whom were female and 49.1% male. A total of 29.6% attended bilingual schools and a further 70.4% came from nonbilingual schools. The participation levels reached the minimum statistical required to be considered representative of this population of learners in the Madrid region.

DATA-COLLECTION INSTRUMENTS

Multiskill English language test

The British Council's *Aptis for Teens* (O'Sullivan et al., 2020), a test specifically designed for 13–17-year-olds, was used to measure English proficiency. It comprised five components: reading, writing, speaking, and listening, plus grammar and vocabulary. The Aptis testing system is based on the sociocognitive model proposed by Weir (2005) and O'Sullivan and Weir (2011) and aligned with the Common European Framework of Reference (CEFR; Council of Europe, 2001). Receptive skills components are automatically scored; meanwhile trained raters mark the speaking and writing responses using an online system. Scores for all components are reported on a 0–50 scale, and CEFR level allocations are given for the four language domains, but not grammar and vocabulary.

Motivational questionnaire

A questionnaire comprising 51 items, delivered in Spanish, captured students' opinions of their schooling and language learning experiences, their language learning motivations, plus a range of socioeconomic indicators, including parental education levels, employment status, and household possessions.

The motivation section of the questionnaire was tailored to the population of 15-year-old Spanish students. There were 32 items in total, divided equally into eight scales:¹

- *Ideal L2 self*—the vision of oneself as a successful language user.
- *Ought-to L2 self*—perceptions of the importance of English based on the opinion of important others in their immediate language learning environment.
- *Language learning experience*—language learning attitudes that can be seen as reflective of language learning experience.
- *Instrumentality*—students' understanding of the utilitarian value of English learning.
- *International orientation*—attitudes toward English as a language which enables learners to communicate with a wide variety of people.
- *English self-concept*—students' perceptions of themselves as learners of English based on their past experiences and social comparison.
- *Parental encouragement*—students' opinions about parental support and their attitudes toward learning English.
- *Motivated learning behavior*—a measure of effort that learners invest in learning English.

The scales were adapted from the motivational questionnaire used by Iwaniec (2014), who also worked with a sample of 15-year-old students, albeit from a different European country (Poland). The exception was the *ought-to L2 self* scale, adapted from Taguchi et al. (2009), which was originally used with a wider age sample (12 to 53 years) in three different contexts: China, Japan, and Iran. The main adaptation involved shortening the scales to four items each, although a small number of other minor changes in wording were made. These were made to adjust the scales to the context studied and to minimize time investment of the participants. To ensure the suitability of the questionnaire for our sample, a pilot study was conducted before the main study.

PROCEDURE

The tests and questionnaires were delivered using an offline-enabled tablet in fully invigilated conditions during a period of one month in March 2017. Full involvement and approval of the Madrid Ministry of Education was obtained prior to conducting this study. Individual participation in the study was contingent on receiving written parental consent.

ANALYSIS

The main data analysis took part in two stages using statistical software Mplus 8 (Muthén & Muthén, 1998–2017). The initial Confirmatory Factor Analysis (CFA) used to confirm the measurement model is summarized in the results that follow and reported in full elsewhere² (Shepherd & Ainsworth, 2017). The analysis novel to this study explored the motivation and proficiency profiles of the participants using LVMM, with the specific model applied being a Factor Mixture Model (FMM).

Confirmatory Factor Analysis

To assess the measurement qualities of the questionnaire, CFA of the current dataset was undertaken. This specified a “measurement model” hypothesizing the eight sets of four observed responses to be indicators of the eight underlying latent variables (please refer to Appendix A for the full path diagram). Whilst this instrument was developed and validated specifically for the use of teenagers learning English in the European Union (Iwaniec, 2015), it had not been applied in the Spanish context before the data-collection exercise described. Verifying this latent structure as a reasonable representation of the data as part of a preliminary assessment was essential before moving forward to the main analysis in the current project, in which the measurement model forms an integral element of the final model, see the following text. To assess whether the hypothesized measurement model achieves this successfully, a number of fit statistics are employed. For the fit indices reported here, it is usually expected that the comparative fit index (CFI) and Tucker–Lewis index (TLI) statistics be *above* 0.9 (or ideally 0.95), and the root mean square error of approximation (RMSEA) *below* 0.05 for good model fit (Byrne, 2012).

Latent Variable Mixture Modeling

The main analysis employed LVMM, a “person-orientated” approach to data modeling, which focuses on characteristics of the individuals from whom the measurements are taken (Bergman & Magnusson, 1997). This analysis identifies unobserved heterogeneity in a population and can be used to divide individuals into subtypes that, as a group, exhibit similar traits or patterns of behavior (Collins & Lanza, 2010).

A range of model structures can be described as LVMMs. The model applied here can be defined as a FMM that falls within category of “hybrid latent variable models” (Muthén, 2007; Muthén & Asparouhov, 2006) because it includes both continuous and categorical latent variables. Our model comprised eight continuous latent factors, each representing a motivational trait. These factors, alongside the five observed test score outcomes, acted as indicators for the latent categorical variable. Classifications could therefore vary with respect to the factor mean for each motivational trait, plus the observed means on the five language skill tests, meaning that each participant was allocated to a latent class in the model based on both motivation and proficiency indicators. A key point to note is that this approach does not assume that the high performers will have high motivation levels, and vice versa, but rather allows the relationship between motivation and proficiency, or indeed between specific measures within each of these, to vary between classes.

Figure 1 plots the relationships modeled, where: C = single latent categorical variable (comprising k categories); m_{1-5} = measured language test scores; f_{1-8} = latent factors representing the motivational traits; u_{1-4} = observations from motivational questionnaire items; and factor loadings are assumed invariant across classes. Important to note here is that the latent class k is hypothesized to explain both the observed test scores (m_{1-5}) and the unobserved motivational traits (f_{1-8}). The results discussion will focus predominantly on how each latent class is characterized with respect to their levels across all these metrics.

The model comprised 37 dependent variables (each of the 5 language test scores, plus the 32 questionnaire responses); 8 continuous latent variables (representing the

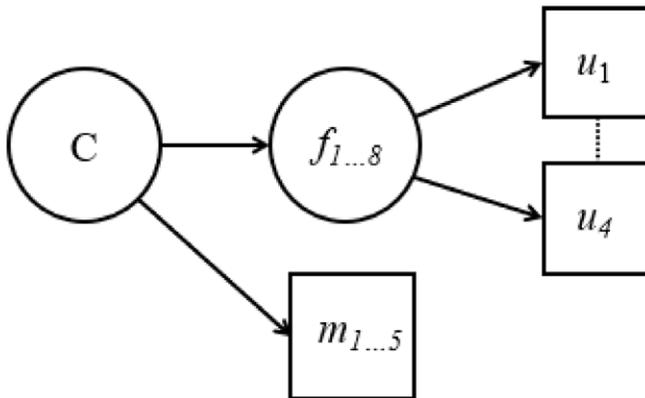


FIGURE 1. Path diagram showing structure of relationships modeled in the current study.

motivational traits); and 1 categorical latent variable (defining the classifications of interest). The model estimated loadings between each question response and its associated continuous latent variable (these were held constant across latent classes); latent means and thresholds for each continuous latent variable (these were allowed to vary between classes, with one class set for reference); and means for each observed test score (these were allowed to vary between classes). The full model specification in Mplus is given in Appendix D.

ANALYTIC AIMS

This investigation first seeks to establish the number of distinct latent classes (k) in the current data with respect both to the participants' English language proficiency and levels of language learning motivation. As with exploratory factor analysis, this is a heuristic technique, and it is not always the case that a definitive answer will arise from the data. In selecting the final number of classes there are statistical indicators that can be drawn upon to aid researchers; however, there is not one single indicator that is accepted for deciding on the number of classes (Nylund et al., 2007). It is best therefore to examine a range of indicators (Masyn, 2013; Nylund et al., 2007). In the modeling exercise reported here, a series of successive models are estimated with the same overall structure (Figure 1) but with an incrementally increasing number of classes specified in the latent categorical variable (C). Model assessment took two stages: (a) examination of statistics to gauge relative model fit and (b) consideration of a range of metrics to judge the quality of class enumeration for the closest fitting models. The evidence accrued from this multistage statistical modeling exercise was then considered against the substantive nature of the class characteristics, as well as the theoretical backdrop provided by previous studies in this field, before establishing the final model. Note there was no explicit consideration of absolute model fit; the standard chi-square statistic can be oversensitive, especially in the case of large datasets, and unfortunately it was not possible to examine the residuals for each response pattern, as recommended by Masyn (2013), as Mplus did not report the output (citing: "the frequency table for the latent class indicator model is too large").³

Under (a), the following statistics provided information about relative fit: Bayesian Information Criterion (BIC; Schwarz, 1978); Consistent Akaike's Information Criterion (CAIC; Bozdogan, 1987); Approximate Weight of Evidence Criterion (AWE; Banfield & Raftery, 1993); and Sample size adjusted BIC (aBIC; Sclove, 1987). Each of these criteria take into account the fit of the model using log-likelihood values whilst penalizing for model complexity (Masyn, 2013). Additionally, as the models with successive numbers of latent classes are nested, likelihood ratio tests can be used to compare two competing models. The LMR (Lo–Mendel–Rubin adjusted likelihood ratio test) compares the improvement in fit between models with adjacent numbers of classes (Lo et al., 2001). This measure provides a p -value, with a significant value suggesting that a greater number of classes are required in the latent variable.

Evaluations of class enumeration under (b) are only considered for models that have already been shown to reach a reasonable level of fit under (a); good class separation in itself does not indicate an overall well-fitting model (Masyn, 2013). *Average class probabilities* (AvePP) provide a measure of the certainty with which model can predict membership of each class (Collins & Lanza, 2010). High classification certainty in the

TABLE 1. Summary of analysis steps and aims

Analysis	Aim
Confirmatory Factor Analysis (CFA)	To confirm the fit of the measurement model specifying the relationship between the 32 observed responses to the motivation questionnaire and their associated latent traits. Note: These results are summarized from a previous study using the same dataset, detailed in Shepherd and Ainsworth (2017).
Factor Mixture Analysis (FMA)	For the main analysis, a Factor Mixture Model (FMM) is built to explore the number of latent classes, or subpopulations, in the data, each characterized by a different motivation-proficiency profile. Considerations are made on the following lines: <ul style="list-style-type: none"> - Relative model fit, - Class enumeration, and - Substantive insights.

model, yields AvePP near 1 with values above 0.7 considered to indicate adequate precision in class assignment (Nagin, 2005). Additionally, as the accuracy of class allocation increases, so do the *odds of correct classification ratio* (OCC; Nagin, 2005). Values above 5 indicate good separation and class assignment (Nagin, 2005). Finally, *entropy* is a standardized index of model-based classification accuracy which ranges from 0 to 1, with values approaching 1 indicating favorable delineation of the classes (Celeux & Soromenho, 1996). The usefulness of entropy lies in highlighting problems with a model, rather than endorsing a particular model choice (Masyn, 2013). It is reported in the current analysis as an indicator of the suitability of a solution, rather than a selection criterion.

Before moving on to discuss the results, a summary of the analytic steps is given in Table 1.

RESULTS

DESCRIPTIVE STATISTICS

The participants displayed a wide range of proficiency levels as evidenced by their overall Apts test performance (see Table 2). The majority of participants were between A2 and B2.

TABLE 2. Distribution of student population for overall test performance by CEFR level (Shepherd & Ainsworth, 2017, p. 44)

CEFR Level	Overall English language test performance		
	Percentage (%)	Standard Error (%)	95% Confidence Interval
C	11.1	0.8	[9.4, 12.7]
B2	22.9	1.2	[20.4, 25.3]
B1	38.5	1.3	[35.9, 41.1]
A2	23.4	1.3	[20.8, 26.0]
A1	4.2	0.6	[2.8, 5.5]

TABLE 3. Scale statistics for the motivation questionnaire

Scale	No. items	Alpha	Scale mean	Scale variance
Ideal L2 self	4	0.91	4.72	0.03
Parental encouragement	4	0.77	4.85	0.08
Self-concept	4	0.84	4.05	0.01
International orientation	4	0.73	5.13	0.17
Ought-to L2 self	4	0.69	3.83	0.47
Language learning experience	4	0.89	4.14	0.19
Motivated behavior	4	0.70	3.80	0.12
Instrumentality	4	0.65	4.40	0.45

Responses to the motivation questionnaire tended, as expected, toward negatively skewed distributions (i.e., leaning toward agree/strongly agree). There was nonetheless a good variation in the options selected, with standard deviations of 1 or more for most items. Alpha statistics ranged from reasonable (>0.6 for *instrumentality*; *ought-to L2 self*) to very good (>0.85 for *ideal L2 self*; *language learning experience*). Table 3 summarizes the descriptive statistics.

MEASUREMENT MODEL

The CFA analysis of the motivational responses reflected the questionnaire design, eight correlated latent traits each with four items loading onto them were tested. The path diagram is shown in Appendix A. This structure displayed reasonable fit statistics (CFI = 0.930; TLI = 0.920; RMSEA = 0.051 [0.049, 0.053]). This analysis was carried out by the authors as part of a previous research project and is reported in full elsewhere (Shepherd & Ainsworth, 2017).⁴ Table 4 summarizes the extent to which each latent variable incorporated in the measurement model accounts for variance in the observed response data, plus the reliability of the constructs (Netemeyer et al., 2003) and coefficient H (Hancock & Mueller, 2001). There is some variation between traits, but overall, it was concluded

TABLE 4. Summary statistics for latent motivation constructs in the measurement model

Scale	No. items	% of variance extracted	Reliability of construct (RC)	Coefficient H [†]
Ideal L2 self	4	71.6%	.91	.91
Parental encouragement	4	48.6%	.79	.81
Self-concept	4	74.9%	.92	.92
International orientation	4	41.9%	.74	.76
Ought-to L2 self	4	51.0%	.81	.81
Language learning experience	4	66.9%	.89	.90
Motivated behavior	4	68.0%	.89	.92
Instrumentality	4	46.9%	.77	.82

[†]Calculations made using online tool from Weiss (2011).

that the questionnaire gave a reasonable reflection of the eight distinct motivational latent traits with all values above 0.7.

FACTOR MIXTURE MODEL

Having established the viability of the measurement model to derive the continuous latent variables representing each of the motivational traits, this section describes the process by which the final FMM was arrived at, with the focus of analysis being on the number of classes incorporated in the latent categorical variable.

Whilst the structure of the model is hypothesized in Figure 1, the exploratory nature of the modeling applied in this study means that no explicit hypothesis is tested regarding the number of latent classes to incorporate in the model. Rather the starting point for this exercise was to hypothesize *some* differences in the motivation and proficiency relationship within the total sample, to which end a two-class model is tested against the one-class model (i.e., a homogeneous population). The number of classes was then increased across successive models until it was clear that adding further classes did not make a statistically significant improvement in the model. In total eight models were run, each incorporating a single latent categorical variable with one to eight classes, respectively. No other element of the model changed between the successive models.

The comparative statistical qualities of the models are described in the following text. It was not possible to derive a measure of overall fit because the software was unable to compute the chi-square test,⁵ however given known sensitivities to large sample size, the chi-square test is likely to have been of limited value. Model evaluation and comparison therefore starts with assessing relative fit for all models, followed by an examination of information about class enumeration for a select range of models. This broadly follows the recommendations given by Masyn (2013). In reaching the final model, statistical findings are evaluated alongside theoretical insights from previous studies.

Relative model fit

Examining the figures for relative model fit provided by the information criteria estimates shown in Table 5, it can be seen that, following a steep drop for all criteria between 1 and 2 categories in the latent variable, the estimate continues to reduce as the number of categories increase, albeit with smaller increments, until a rise between seven and eight

TABLE 5. Model comparison: Information criteria

Model	BIC	SSBIC	CAIC	AWE
1-class	230594.81	229978.48	230788.81	232628.01
2-class	216199.64	215538.83	216407.64	218379.56
3-class	211577.00	210871.72	211799.00	213903.65
4-class	208909.93	208160.17	209145.93	211383.31
5-class	207349.06	206554.83	207599.06	209969.16
6-class	206188.22	205349.51	206452.22	208955.05
7-class	205540.49	204657.30	205818.49	208454.05
8-class	205598.33	204670.67	205890.33	208658.62

categories. Table 6 shows the Approximate Correct Model Probabilities (cmP_A; Masyn, 2013) based on the Schwarz Information Criterion (SIC [-0.5 BIC]; Schwarz, 1978) that gives a comparative probability that each model is the correct model based on all models under consideration. These figures suggest that the 7-class model has the edge over models containing other numbers of latent classes, though the distinction is only marginal. This initial information would suggest that a latent variable incorporating up to seven categories would be statistically beneficial to the model, however from five categories onward, marginal gains become negligible (see Figure 2).

TABLE 6. Approximate Correct Model Probabilities, comparing models including 2–8 classes

Number of classes (<i>k</i>)	Approximate Correct Model Probability (cmP_A)
2	0.00
3	0.09
4	0.14
5	0.17
6	0.19
7	0.21
8	0.20

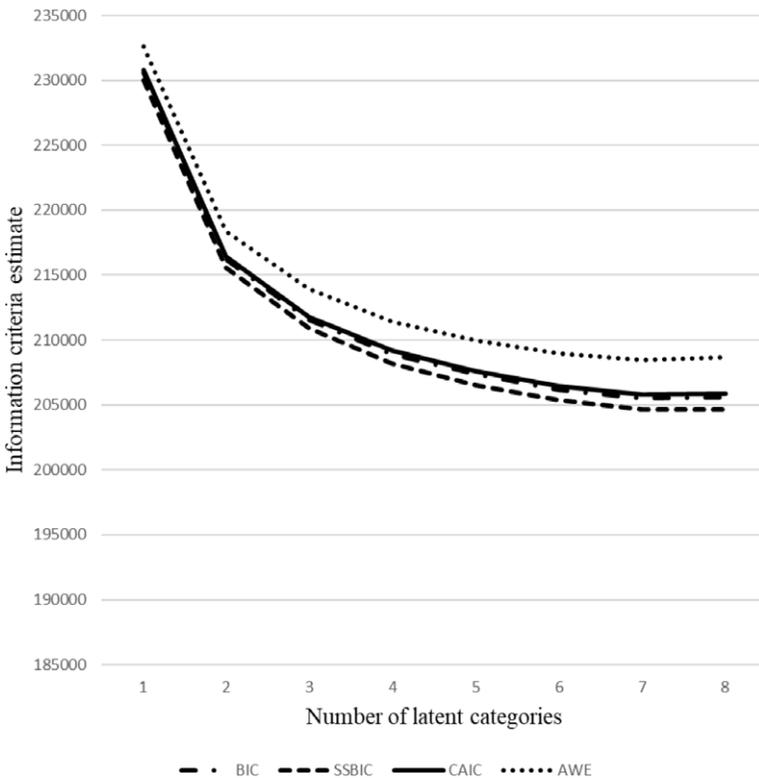


FIGURE 2. Elbow plot mapping information criteria estimates for successive models.

Although the IC-based evidence presented in the preceding text indicates that model improvements are derived up to the inclusion of seven categories in the latent variable, the *p*-value associated with the LMR becomes nonsignificant with the addition of the sixth category. See Table 7 for full details.

Class enumeration

In considering class enumeration across a restricted range of models, the relative entropy indicates that there are no problematic classifications in any of the models, as all are above .9. However, this averaged figure can mask latent class assignment error for specific individuals in the data. In comparing the AvePP for each model shown in Table 8, it is notable that the model incorporating the 5-class latent variable sees the most even dispersal of probabilities across classes, each clustered around .95. There is greater variation in classification probability for each of the other models listed. This pattern is also reflected in the OCC estimates shown in Table 9.

Overall, having considered the evidence presented in the preceding text, the statistical indicators suggest that the latent categorical variable should incorporate *at least* five latent classes, with potential gains from including as many as seven. To make the final decision on class number, estimates for each of the three models were compared, and the relative value of the substantive insights considered.

With respect to the language skills profile, the 5-, 6-, and 7-class models all revealed classes of students that fell into one of two strata of achievement, higher and lower. Without having the space to go into full details, it was determined that the 5-class model

TABLE 7. Model comparisons: Likelihood ratio tests

Number of classes (<i>k</i>)	Parameter difference	LMR adjusted LRT	
		Adjusted 2LL	<i>p</i> value: <i>k</i> versus <i>k</i> -1 classes
2	14	14449	<.001
3	14	4683	<.001
4	14	2746	<.001
5	14	1650	0.002
6	14	1254	0.700
7	14	745	0.299

TABLE 8. Relative entropy and average latent class probabilities for most likely class membership for models incorporating 4–7 latent categories

Number of classes	Relative entropy	AvePP for each category in model						
		1	2	3	4	5	6	7
5	0.930	0.948	0.956	0.958	0.957	0.955	–	–
6	0.917	0.952	0.927	0.947	0.952	0.916	0.954	–
7	0.911	0.970	0.914	0.928	0.923	0.921	0.925	0.960

TABLE 9. Odds of correct classification ratios (OCC) for models incorporating 4–7 latent categories

No. categories	OCC estimates for each category in model						
	1	2	3	4	5	6	7
5	107.05	67.45	205.67	73.29	55.10	–	–
6	183.44	48.48	105.88	77.84	38.04	145.64	–
7	319.38	54.13	51.30	92.19	68.77	56.65	206.02

reflected a core set of relationships between proficiency and motivation, with the 6- and 7-class models adding shades of distinction to this with respect to relationships with individual motivation traits. In this respect, the additional classes in the 6- and 7-class models were understood not to reflect substantively distinct categorizations of learners. The decision was taken therefore to move forward with the 5-class model. Incorporating five classes in the categorical latent variable reflects the class enumeration settled on in the previous motivation studies discussed in the preceding text; it also represents the most parsimonious of the statistically acceptable options in the current modeling exercise, thus avoiding overly convoluting the interpretation. The discussion of the results in the text that follows takes a closer look at the pattern of language test scores and motivation factor scores for students in each class.

SUBSTANTIVE FINDINGS

The final model incorporating five classes in the latent categorical variable was created. To reach a stable solution (McLachlan & Peel, 2000; Muthén & Muthén, 1998–2017) the number of random starts was increased to 2,000 (from the software default of 20). With this number of starting values, the best log-likelihood was replicated 30 times. This gave a good indication that localized solutions were avoided, important given the complexity of the model under consideration. The estimates from this model are given in the following text. Table 10 shows that there is a reasonable representation across the sample of all five classes, with class 1 having the fewest members and class 4 the greatest.

Mean scores across each of the language subskill tests and for each of the motivational traits are given in Tables 11 and 13, with Figures 3 and 4 depicting these estimates

TABLE 10. Class counts and proportions

Latent class	Number of students	Proportion
1	177	0.10
2	432	0.24
3	258	0.15
4	493	0.28
5	413	0.23

TABLE 11. Mean observed scores for language skill tests by class

Measure	Class 1	Class 2	Class 3	Class 4	Class 5
Grammar and Vocabulary	16.38	18.03	19.39	29.99	33.83
Listening	26.40	29.20	31.68	41.43	43.79
Reading	18.39	20.31	21.91	38.29	41.61
Speaking	14.34	20.45	25.26	36.25	38.80
Writing	16.79	22.57	25.26	38.37	40.43

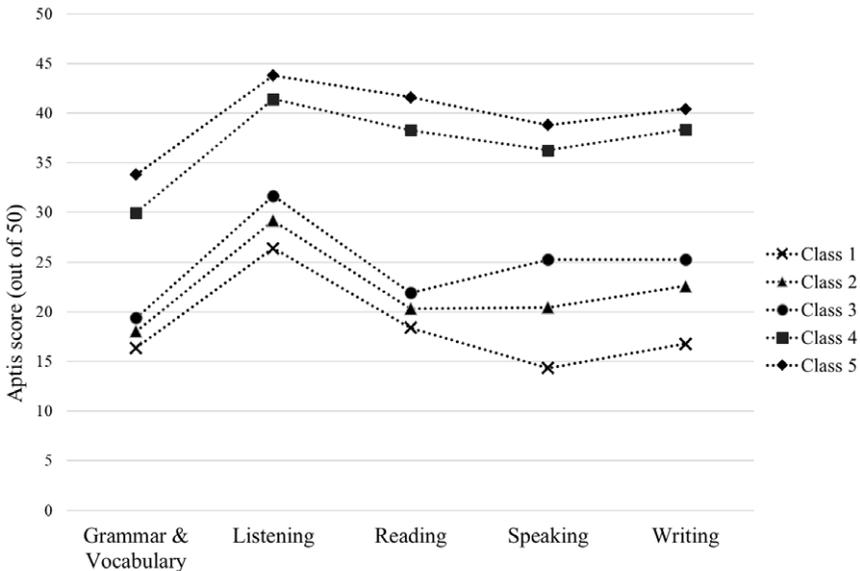


FIGURE 3. Mean Aptis scale scores by class.

visually. The mean scores for the language skill tests are out of a possible total⁶ of 50. Figure 3 shows two strata of achievement, with a reasonably wide separation between classes 4 and 5, and classes 1, 2, and 3. The distinction in performances for classes 1, 2, and 3 is marginally greater for productive than for receptive skills. A Kruskal–Wallis test indicates that the overall differences in L2 proficiency between classes is significant ($p < .001$), with Mann–Whitney U tests indicating significant differences between each pair of classes for each skill area (nonparametric tests were used owing to violation of the normality assumption). However, whilst these mean score differences are significantly different, with respect to overall English language ability, it should be acknowledged that the abilities represented in each class encompass a reasonably large ability range as shown by the overall CEFR level by class in Table 12. This also plays out at the individual skill level, please see Appendix B for details.

Motivational trait levels are given as unweighted factor scores derived from the model, calculated in Mplus using *expected a posteriori* (EAP) estimation. It should be also noted

TABLE 12. Count of overall CEFR level by class

CEFR	Class 1	Class 2	Class 3	Class 4	Class 5	Total
A1	34	27	5	0	0	66
A2	112	193	81	0	1	387
B1	31	208	169	183	67	658
B2	0	4	3	222	198	427
C	0	0	0	88	147	235

TABLE 13. Mean factor scores for motivational traits by class

Measure	Class 1	Class 2	Class 3	Class 4	Class 5
F1—International orientation	-2.95	-1.67	-0.38	-1.56	0.000 ¹
F2—Parental encouragement	-1.22	-0.76	0.06	-0.62	0.000 ¹
F3—Self-concept	-6.22	-4.38	-2.28	-2.24	0.000 ¹
F4—Ought-to L2 self	-1.13	-0.65	0.03	-0.56	0.000 ¹
F5—Language learning experience	-5.83	-3.39	-0.79	-2.79	0.000 ¹
F6—Motivated behavior	-4.21	-2.32	-0.48	-2.12	0.000 ¹
F7—Instrumentality	-2.90	-1.89	-0.70	-1.06	0.000 ¹
F8—Ideal L2 self	-6.44	-4.26	-1.59	-2.52	0.000 ¹

¹Reference category.

that the scale of the factor scores bears no relationship to the original scale of the response options, but rather is determined within the statistical model. The mean of the factors for class 5 are fixed to zero for identification purposes and to set the metric of the factor distribution (Clark et al., 2013). Factor scores for each trait for the other classes of students should therefore be interpreted in relation to the baseline provided by class 5, with the chart in Figure 4 to be understood as showing the points of departure of the first four classes from the reference category, class 5. Note this does not mean that class 5 had uniform motivation levels because the estimation of factor scores for each trait are independent of each other.

Figure 4 emphasizes differences in the relative degree of differentiation between classes across each trait. It is notable that much greater separation occurs between classes for the internalized traits of motivation: ideal L2-self, self-concept, and language learning experience. The traits with the closest estimates between groups are the external motivating factors of ought-to L2 self and parental encouragement.

Nonsignificant differences between trait levels are only observed in the following:

- International orientation: class 2 and class 4 ($p = 0.613$)
- Parental encouragement: class 2 and class 4 ($p = 0.389$); class 3 and class 5 ($p = 0.520$)
- Self-concept: class 3 and class 4 ($p = 0.866$)
- Ought-to L2 self: class 2 and class 4 ($p = 0.556$); class 3 and class 5 ($p = 0.691$)
- Motivated behavior: class 2 and class 4 ($p = 0.423$)

The remaining factor score means each varied significantly at the 5% level between classes for each trait. A full report of the t-statistics and p values is given in Appendix C.

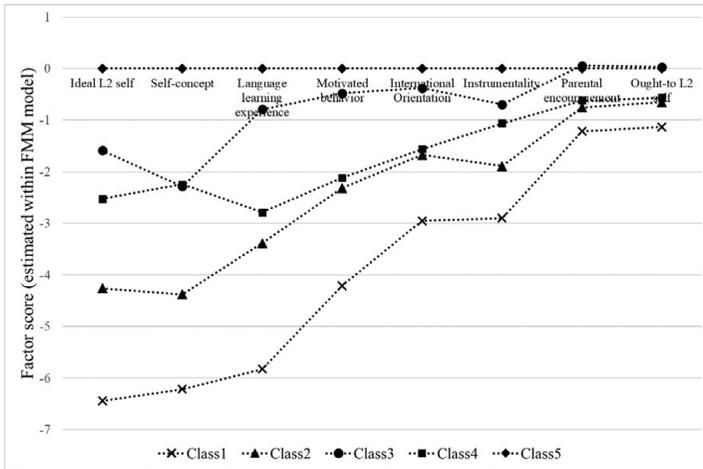


FIGURE 4. Mean factor scores for motivational traits by class.

A key point to note from Figures 3 and 4 is that whilst the highest and lowest performing classes report the highest and lowest levels of motivation across all traits respectively, the interim classes do not follow a pattern linking proficiency to motivation in a straightforward fashion. Conspicuously, class 3, the higher performers in the lower strata, have the second highest levels of motivation, followed by high achievers in class 4. This switch between classes 3 and 4 would not have been detected in an analytic approach that assumed homogeneity in the motivation/proficiency relationship across the full sample.

DISCUSSION

This is the first study to incorporate L2 skills measured by a robustly validated test of English language into profiles of language learning motivation. As well as demonstrating the motivational traits that separate students of differing proficiency levels to the greatest extent, the study can be viewed as a solid addition to the body of evidence providing support to the L2MSS. The value of using profile analysis to gain insights into the relationship between language learning motivation and proficiency is established under rigorous methodological conditions. In particular, by questioning the assumption of linearity in the relationship between motivation and proficiency variables made in previous correlational studies, the person-centered approach employed here reveals that this does not apply in all cases. A distinct and sizable group of students were shown to violate this pattern (class 3; 15% of students). In a variable-centered correlation analysis the contribution of these class 3 learners would have weakened the relationship but would not have refuted the overall picture. This highlights the sensitivity of the profiling methodology to nuances in the data that may have been overlooked.

Each latent class in the statistical model defines a group of students who exhibit commonalities with respect to the motivation and proficiency scales, each with

straightforwardly interpretable characteristics. To this end we have been able to understand the five classes in the following broad terms: class 5 as *high-flying*; class 4 as *coasting*; class 3 as *aspirational*; class 2 as *ambivalent*; and class 1 as *uninvested*. Further details of this taxonomy are given in the following text. It should be highlighted for clarity that these classes of learners have been identified based on their pattern of responses, rather than hypothesized *a priori*. Employing this profiling approach opens options for understanding more clearly how and where policy intervention would be best targeted, by fully engaging with the distinguishing features of the learners in each group.

LEARNER PROFILES

Five motivation-proficiency profiles were identified amongst the sample of teenage language learners in Madrid. The results bear some similarities to studies profiling language learning motivations (Csizér & Dörnyei, 2005; Kangasvieri, 2017; Papi & Teimouri, 2014; Yun et al. 2018), in which multiple motivational profiles were identified. The current study therefore provides support for the perspective that learner motivation can be better understood in terms of a range of distinct motivational profiles, with the additional value of coalescing this approach with information about L2 skill levels. In doing so, we were able to show that increases in motivational trait scores are not always matched with a straightforward increase in L2 proficiency. In this section we discuss the classifications that emerge under this person-centered approach, and the characteristics of learners within each.

Students in class 1 unflinchingly scored lowest on all the measures of both proficiency and motivation. As a result, this smallest class of students (10% of the sample; Table 5) can be referred to as “uninvested” learners. Students in class 2 (24% of the sample) have slightly higher proficiency than uninvested students and displayed somewhat higher levels of motivation. Because both their motivation and proficiency are still on the low side, these learners can be referred to as generally “ambivalent.” These uninvested and ambivalent classes of students can be contrasted against students in class 5, the “high-flyers.” High-flyers (23% of the sample) are characterized by the highest levels of proficiency of all classes and declare highest average motivation on six out of eight motivational variables.

To this extent, these findings are similar to those of the other studies (Csizér & Dörnyei, 2005; Kangasvieri, 2017; Papi & Teimouri, 2014; Yun et al. 2018) who also identified a clear separation between a group of very low and very highly motivated learners across all traits. More specifically, the high-flyers bear resemblance to the Yun et al.’s “thrivers,” with their exceptionally strong motivation, and the low motivation of ambivalent and uninvested learners are similar to the “dependent” profile from Yun et al.’s study. However, given some salient differences in the nature of the traits between the current study and that of Yun et al., it is not possible to draw direct comparisons.

Learners in class 4 (just under 28% of the sample) achieved relatively high levels of proficiency yet displayed only moderate motivation levels. On four of the motivation traits (international orientation, parental encouragement, ought-to L2 self, and motivated behavior), mean factor scores of students in this class are not significantly different from those of the ambivalent (class 2) class of learners. Additionally, although statistically significant, the numerical differences between these two classes for language learning experience and instrumentality are not large. Class 4, therefore, despite having good

proficiency levels in English, could be described as not particularly engaged with the benefits of learning English, and perhaps “coasting” along. This bears resemblance to the finding reported by Hessel (2017), who reported that upon reaching a relatively high level of productive proficiency, some learners may lose the motivational impetus for further study. In contrast to class 2 however, class 4 students tend to hold more robust visions of themselves as language learners, perhaps unsurprisingly given their higher achievement levels.

Students in class 3 are the most “aspirational” group of learners in the sample. They have, on average, slightly higher proficiency than those in classes 1 and 2, yet they exhibit comparatively far stronger motivation to learn English. In comparison to the high-achieving class 4 meanwhile, their motivation levels were higher on all but one of the eight traits. The exception being the English self-concept scale, for which there was no significant difference between these two classes. It is also noteworthy that the aspirational learners in class 3 also reported the highest average levels of parental encouragement and ought-to L2 self of all classes.

Whilst the levels of motivation and proficiency of students from classes 1, 2, 4, and 5 largely rise in tandem, suggesting a positive relationship between motivation and proficiency, the emergence of the aspirational class 3 breaks this trend and adds complexity to the picture. One explanation could be that these students have a strong desire to learn English, but not the same access to resources to realize this vision. However, the fact that students in class 3 were attendant at a full range of schools in the sample indicates that this may be an overly simplistic interpretation. Indeed, it could be that these students have experienced a recent surge in motivation, the effects of which can be only accounted for in the future. In some respects, this group of students resembles the Indonesian learners of English interviewed by Lamb (2013), who were highly motivated, but of limited proficiency, with their future vision being almost “dreamlike” in nature. This temporal dimension represents a key complication in modeling the relationship between motivation and proficiency in a cross-sectional study. Where the proficiency levels represent the current ability of the participants in the L2 language, the motivational traits encompass a wider temporal domain. The situation with the learners in class 3 might be that they will go on to make much stronger progress than their peers in classes 1 and 2, but it is not possible to say from the current findings. The relatively greater prowess in productive skills compared to other classes in the lower strata of ability (classes 1 and 2) might be indicative of this incipient progression. For these aspirational learners, the relatively higher levels of productive skill performance, particularly speaking, is perhaps already reflective of this desire to use English to communicate, as expressed in their strong international orientation and understanding of the instrumental value of English.

THE RELATIONSHIP BETWEEN MOTIVATION AND PROFICIENCY

This section examines more closely how the findings compare to existing understanding of the motivation-proficiency relationship as presented in existing SLA studies. As observed above with respects to Figures 3 and 4, certain motivational traits separate classes of students to a larger extent than others. The traits displaying the most exaggerated gaps between classes of the learners with the highest and lowest English language

proficiency are ideal L2 self, English self-concept, and language learning experience. This pattern indicates that these motivational traits play a key role in differentiating students at differing proficiency levels, although the profiling approach has helped us to emphasize why these relationships are not best expressed as straightforward correlations (see discussion regarding aspirational learners in class 3 in the preceding text). The findings from this study are broadly consistent with those reported by Kangasvieri (2017) who also found that variables of self-concept and ideal L2 self to be robust differentiators between learners, and Al-Hoorie (2018) whose meta-analysis pointed to the positive role played by ideal L2 self and language learning experience in developing proficiency. In addition, Papi and Teimouri (2014) reported the same for ideal L2 self and language learning experience.

The finding that language learning experience acts as an important differentiator between high and low proficiency students is in line with those of Lamb (2012), who showed that learning experience in school could explain variance in the results of a proficiency test. This is the only other study to report links between language learning experience and proficiency. However, more broadly, previous literature has pointed to a close relationship between language learning experience and measures of effort investment (Csizér & Kormos, 2009; Ryan, 2009). Whilst this is not a direct association, such observations do imply that positive language learning experience is associated with a mindset that will engender heightened proficiency.

A similar second-order influence on proficiency can be found in the literature with respect to the relationship between ideal L2 self and students' reported effort investment (see for example Iwaniec [2014]; Csizér and Kormos [2009]). These findings lend some credence to the more direct relationship between ideal L2 self and L2 proficiency reported here. However, this relationship is not clear-cut, for although our findings were in accordance with results reported by Saito et al. (2019), these findings contradict those of an earlier study from Moskovsky et al. (2016). Such discrepancies might well be ascribed to the differences between the contexts, or measures applied. The robust sampling methodology and comprehensive range of language proficiency measures incorporated in the current study make it a good baseline for future comparisons.

One key insight to be gleaned from the findings of the current study is that the ought-to L2 self does not appear to play a crucial role in determining language learning proficiency amongst this learner population, as there were no marked variations in the level of this variable between the five classes. A similar finding was made for the other externalized motivation trait: parental encouragement. It appears that the perception of the importance of English as communicated by important others has limited motivational properties for teenage learners in Madrid. This finding is in accordance with previous studies, most of which agreed about a lack of clear relationship between the ought-to L2 self and attainment (Al-Hoorie, 2018; Lamb, 2012; Moskovsky et al., 2016; Saito et al., 2019; for exceptions see Papi et al. [2019], who found clear links between ought-to L2 self/own and motivated behavior in the ESL context of the United States).

Interestingly, the findings point to a weaker link between the measures of effort investment, as conveyed by the motivated behavior scale, in affecting English language proficiency, than the ideal L2 self and proficiency. This perhaps can be attributed to the

“hybrid attributes” (Dörnyei & Ushioda, 2011) of the ideal L2 self, which is not only affective by nature but is also likely to have a cognitive component. An example of this is that the learner makes judgments on how plausible achieving their vision is. It is also possible that the comparatively weaker role for effort investment as reported in this study relates to a possible interference of other variables on the quality of effort that the Spanish students invest in language learning, such as language learning aptitude, the use of language learning strategies, or the degree of autonomy and self-regulation, for which the current model does not account.

Finally, the two language learning goals investigated in this study, international orientation and instrumentality, are shown to have a moderate potential to differentiate between students of varying proficiency. This finding is unsurprising, as the two goals are long term and need to be translated into a series of smaller goals to sustain motivation (Ford, 1992). Although their motivational properties are not as strong as when the goals are integrated into a robust vision of themselves in the future, taken together, our finding suggests that they are more likely to contribute to proficiency than external pressures from parents and society.

CONCLUSION

This study identified five heterogeneous classes of learners amongst a representative sample of teenage language learners in the Madrid region. From the “ambivalent,” “uninvested,” and “coasting” students to the “aspirational” learners and the “high-flyers,” these students displayed a wide range of attitudes toward English language learning reflected in equally varied levels of proficiency. The value of taking a person-centered profiling approach to the analysis was that it was able to demonstrate that whilst there is undoubtedly a connection between motivation and proficiency, it is not a straightforward correlation.

The study contributes to a body of research that aims at substantiating Dörnyei’s (2005) L2MSS. The majority of such studies have examined the relationship between the three main constructs of the L2MSS—ideal L2 self, ought-to L2 self, language learning experience—and measures of effort investment. The current study extends this by investigating links with proficiency for these three constructs and proficiency, plus motivated learning behavior. Similar to the previous research, our findings confirm the important role of the ideal L2 self and the learning environment in motivating English language learning. In contrast, our findings point to the limited potential of the ought-to L2 self in motivating 15-year-old Spanish students to learn English. At the same time, we have shown that employing additional motivational constructs can enrich our understanding of how motivation and proficiency are related. Of most importance here is the English self-concept, which differentiates clearly between participants of different proficiency levels. Traits indicating the extent to which learners identify with language learning goals (international orientation and instrumentality) provided additional insight, spanning a middle ground between the strong role of internalized traits and the weak role identified externalized traits. In contrast, including the parental encouragement scale strengthened the conclusion that external pressures with which the learners do not identify are of little motivational value for the Spanish teenagers.

This study makes a key contribution to research into SLA motivation by demonstrating the need to move away from the assumption of a uniform relationship between motivation and proficiency for all groups of learners. The person-centered approach side-stepped this assumption, thus providing the opportunity for a more intricate picture of the dynamic to emerge. This lays the groundwork for future research into different classes of language learners based on their motivation and proficiency. However, whilst there is a consensus that different motivational profiles of learners exist, little is known about why these profiles develop. In particular, investigations that account for the existence of “aspirational” learners who are highly motivated, yet with relatively low proficiency, are necessary. Similarly, research into “uninvested” learners could help us better understand what makes learners relatively successful, despite modest levels of motivation.

Motivation is one of many variables that interact with others in a dynamic way during the process of language learning. One limitation of this study is that other individual differences (e.g., personality) that could potentially have shed more light on the results were not included in the analysis. Another source of weakness in this study is the fact that the measures were taken at a single moment in time. Although it can be broadly assumed that the participants’ proficiency was increasing with time, the participants’ might have experienced fluctuations in their levels of motivation. A longitudinal design would be helpful in a more intricate unraveling of the relationship between motivation and proficiency. Additionally, the two-stage sampling approach was not explicitly accounted for in the analysis. Because the sample was clustered within a specified number of schools, this may have reduced variation in the dataset compared to a straightforward random sample, moving away from complete independence of observations between certain clusters of students. The shared educational experience of the students attending the same school may have engendered commonalities in motivation and proficiency relations. However, the school sample saw participation of 24% of the public schools within the Madrid region (169/707 schools), with between 8–12 students sampled at each school (mean number of students at each school: 10.5) (Shepherd and Ainsworth, 2017). Therefore, given this school coverage, and the already complex nature of the statistical model applied, it was decided to move forward with treating the data as a simple random sample for the purpose of this exploratory analysis. It should also be noted that the sampling approach applied is much less open to bias than smaller convenience samples often used for studies in this field.

The process of exploring motivation-proficiency profiles described in this article provides a new approach to unpicking this complex relationship. The value of this level of insight lies in the possibilities to highlight areas where educational and policy interventions could be of real benefit to learners; for example, it could be suggested that the aspirational class 3 could be most likely to benefit from positive interventions. This group is not an observed group of students classified by gender, school type, or social background, they are not picked out *a priori* for comparison, rather they are defined by the analysis. In taking the analytic approach that assumes commonalities in responses to language education not to necessarily be defined by such visible markers, this study shows how we as educationalists can break out of the usual comparative framework and make more astute appraisals of the situation.

NOTES

¹The materials are available at <https://www.iris-database.org/iris/app/home/detail?id=york:939629>.

²Please note that the motivation questionnaire analysis described in Shepherd and Ainsworth (2017) was conducted using the same dataset as used for the current study.

³It should be noted that the model fit was large and complex, requiring approximately 24 hours computation time in Mplus.

⁴In the original analysis, improvements to the model were achieved by incorporating error covariances between four of the observed variables. However, for the purposes of moving forward to building this series of latent variables into the successive FFMs it was decided that the fit achieved without correlating the error covariances was acceptable. The rationale for this was not to overly convolute an already complex model.

⁵Returning the message for all models: “The chi-square test cannot be computed because the frequency table for the latent class indicator model part is too large.”

⁶Note that the calibration of each of these tests is different, so the score in one skill area should not be compared directly to that of another area, i.e. a higher numerical listening than writing score does not necessarily equate to greater skill in listening than writing.

REFERENCES

- Anisfield, M., & Lambert, W. E. (1961). Social and psychological variables in learning Hebrew. *Journal of Abnormal and Social Psychology*, 63, 524–529. <https://doi.org/10.1037/h0043576>
- Al-Hoorie, A. H. (2018). The L2 motivational self-system: a meta-analysis. *Studies in Second Language Learning & Teaching*, 8, 721–754. <https://doi.org/10.14746/ssl.t.2018.8.4.2>
- Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49, 803–821. <https://doi.org/10.2307/2532201>
- Bergman, L. R., & Magnusson, D. (1997). A person-orientated approach in research on developmental psychopathology. *Development and Psychopathology*, 9, 291–319. <https://doi.org/10.1017/S095457949700206X>
- Bozdogan, H. (1987). Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345–370. <https://doi.org/10.1007/BF02294361>
- Byrne, B. M. (2012). *Structural Equation Modelling with Mplus: Basic concepts, applications, and programming*. Routledge.
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195–212. <https://doi.org/10.1007/bf01246098>
- Clark, S. L., Muthén, B., Kaprio, J., D’Onofrio, B. M., Viken, R., and Rose, R. J. (2013) Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling*, 20, 681–703. <https://doi.org/10.1080/10705511.2013.824786>
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioural, and health sciences*. Wiley.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- Csizér, K., & Dörnyei, Z. (2005). The internal structure of language learning motivation and its relationship with language choice and learning effort. *The Modern Language Journal*, 89, 19–36. <https://doi.org/10.1111/j.0026-7902.2005.00263.x>
- Csizér, K., & Kormos, J. (2009). Learning experiences, selves and motivated learning behaviour: A comparative analysis of structural models for Hungarian secondary and university learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 98–119). Multilingual Matters.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum Associates.
- Dörnyei, Z., & Chan, L. (2013). Motivation and vision : An analysis of future L2 self images, sensory styles , and imagery capacity across two target languages. *Language Learning*, 63, 437–462. <https://doi.org/10.1111/lang.12005>
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Pearson Education Limited.

- Feng, L., & Papi, M. (2020). Persistence in language learning: The role of grit and future self-guides. *Learning and Individual Differences, 81*, 101904. <https://doi.org/10.1016/j.lindif.2020.101904>
- Ford, M. (1992). *Motivating humans: Goals, emotions, and personal agency beliefs*. Sage Publications.
- Gardner, R. C. (1960). *Motivational variables in second-language acquisition*. (Unpublished PhD Thesis). Montreal: McGill University.
- Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second language acquisition. *Canadian Journal of Behavioural Psychology, 13*, 266–272.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Newbury House Publishers.
- Gardner, R. C., Tremblay, P. F., & Masgoret, A.-M. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal, 81*, 344–362. <https://doi.org/10.1111/j.1540-4781.1997.tb05495.x>
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.) *Structural equation modeling: Present and future—A Festschrift in honor of Karl Joreskog* (pp. 195–216). Scientific Software International.
- Hessel, G. (2017) A new take on individual differences in L2 proficiency gain during study abroad. *System, 66*, 39–55. <https://doi.org/10.1016/j.system.2017.03.004>
- Iwaniec, J. (2014). Self-constructs in language learning: What is their role in self-regulation? In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 189–205). Multilingual Matters.
- Iwaniec, J. (2015). *Motivation to learn English of Polish gymnasium pupils*. (PhD thesis). Lancaster University.
- Kangasvieri, T. (2017). L2 motivation in focus: The case of Finnish comprehensive school students. *The Language Learning Journal, 47*, 188–203. <https://doi.org/10.1080/09571736.2016.1258719>
- Kim, T.-Y., & Kim, Y.-K. (2014). EFL students' L2 motivational self system and self-regulation: Focusing on elementary and junior high school students in Korea. In K. Csizér & M. Magid (Eds.), *The impact of self-concept on language learning* (pp. 87–107). Multilingual Matters.
- Kim, Y.-K., & Kim, T.-Y. (2011). The effect of Korean secondary school students' perceptual learning styles and ideal L2 self on motivated L2 behavior and English proficiency. *Korean Journal of English Language and Linguistics, 11*, 21–42. <https://doi.org/10.15738/kjell.11.1.201103.21>
- Lamb, M. (2012). A self system perspective on young adolescents' <https://doi.org/10.1111/j.1467-9922.2012.00719.x>
- Lamb, M. (2013). "Your mum and dad can't teach you!": Constraints on agency among rural learners of English in the developing world. *Journal of Multilingual and Multicultural Development, 34*, 14–29. <https://doi.org/10.1080/01434632.2012.697467>
- Lambert, W. E., Gardner, R. C., Barik, H. C., & Tunstall, K. (1972). Attitudinal and cognitive aspects of intensive study of a second language. In R. C. Gardner & W. E. Lambert (Eds.), *Attitudes and motivation in second-language learning*. Newbury House Publishers.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767–778. <https://doi.org/10.1093/biomet/88.3.767>
- Marcoulides, G. A., & Heck, R. H. (2013). Mixture models in education. In T. Teo (Ed.), *Handbook of quantitative methods for educational research* (pp. 347–366). SensePublishers.
- Masyn, K. E. (2013) Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods*. Vol. 2: Statistical Analysis. Oxford University Press.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley.
- Michaelides, M., Brown, G., Eklöf, H., & Papanastasiou, E. (2019). *Motivational profiles in TIMSS Mathematics: Exploring student clusters across countries and time*. Springer Open. <https://www.springer.com/gp/book/9783030261825>
- Moskovsky, C., Assulaïmani, T., Racheva, S., & Harkins, J. (2016). The L2 Motivational Self System and L2 achievement: A study of Saudi EFL learners. *The Modern Language Journal, 100*, 641–654. <https://doi.org/10.1111/modl.12340>
- Muthén, B. (2007) Latent variable hybrids: Overview of old and new Models. In G. R. Hancock and K. M. Samuelson (Eds.), *Advances in latent variable mixture models*. Information Age
- Muthén, B. O., & Asparouhov, T. (2006) Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors, 31*, 1050–1066. <https://doi.org/10.1016/j.addbeh.2006.03.026>
- Muthén, L. K., & Muthén, B.O. (1998–2017). *Mplus User's Guide*. 8th ed. Muthén & Muthén.
- Nagin, D. S. (2005). *Group-based modeling of development*. Harvard University Press.

- Netemeyer, R., et al. (2003). *Scaling procedures: Issues and applications*. SAGE. <https://doi.org/10.4135/9781412985772.n3>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 535–569. doi:10.1080/107051070151575396
- O'Sullivan, B., & Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language testing: Theories and practices* (pp. 13–32). Palgrave Macmillan.
- O'Sullivan, B., Dunlea, J., Spiby, R., Westbrook, C., & Dunn, K. (2020). Apts General Technical Manual, Version 2.2. TR/2020/001. British Council, London. <https://www.britishcouncil.org/exam/apts/research/publications/technical/general-technical-manual-version-2-2>
- Papi, M., Bondarenko, A. V., Mansouri, S., Feng, L., & Jiang, C. (2019). Rethinking L2 motivation research: The 2 × 2 model of L2 self-guides. *Studies in Second Language Acquisition*, 41, 337–361. <https://doi.org/10.1017/S0272263118000153>
- Papi, M., & Khajavy, G. H. (2021). Motivational mechanisms underlying second language achievement: A regulatory focus perspective. *Language Learning*, 71, 537–572. <https://doi.org/10.1111/lang.12443>
- Papi, M., & Teimouri, Y. (2014). Language learner motivational types: A cluster analysis study. *Language Learning*, 64, 493–525. <https://doi.org/10.1111/lang.12065>
- Peal, E., & Lambert, W. E. (1972). The relation of bilingualism to intelligence. In R. C. Gardner & W. E. Lambert (Eds.), *Attitudes and motivation in second language learning* (pp. 228–245). Newbury House.
- Ryan, S. (2009). Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 120–144). Multilingual Matters.
- Saito, K., Dewaele, J., Abe, M., & In'nami, Y. (2019). Motivation, emotion, learning experience and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, 68, 1–35. <https://doi.org/10.1111/lang.12297>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464. <https://www.jstor.org/stable/2958889>
- Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333–343. <https://doi.org/10.1007/BF02294360>
- Shavelson, R., Hubner, J., & Stanton, G. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46, 407–441. <https://doi.org/10.3102/00346543046003407>
- Shepherd, E., & Ainsworth, V. (2017). *English impact: An evaluation of English language usage capability*. <https://www.britishcouncil.es/sites/default/files/british-council-english-impact-report-madrid-web-opt.pdf>
- Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 66–97). Multilingual Matters.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.
- Weiss, B. A. (2011). *Reliability and validity calculator for latent variables* [Computer software]. <https://blogs.gwu.edu/weissba/teaching/calculators/reliability-validity-for-latent-variables-calculator/>
- Yashima, T. (2000). Orientations and motivation in foreign language learning: A study of Japanese college students. *JACET Bulletin*, 31, 121–133.
- Yun, Y., Hiver, P., & Al-Hoorie, A.H. (2018). Academic buoyancy: Exploring learners' everyday resilience in the language classroom. *Studies in Second Language Acquisition*, 40, 805–830. <https://doi.org/10.1017/S0272263118000037>.

APPENDIX A

MEASUREMENT MODEL FOR MOTIVATIONAL TRAITS

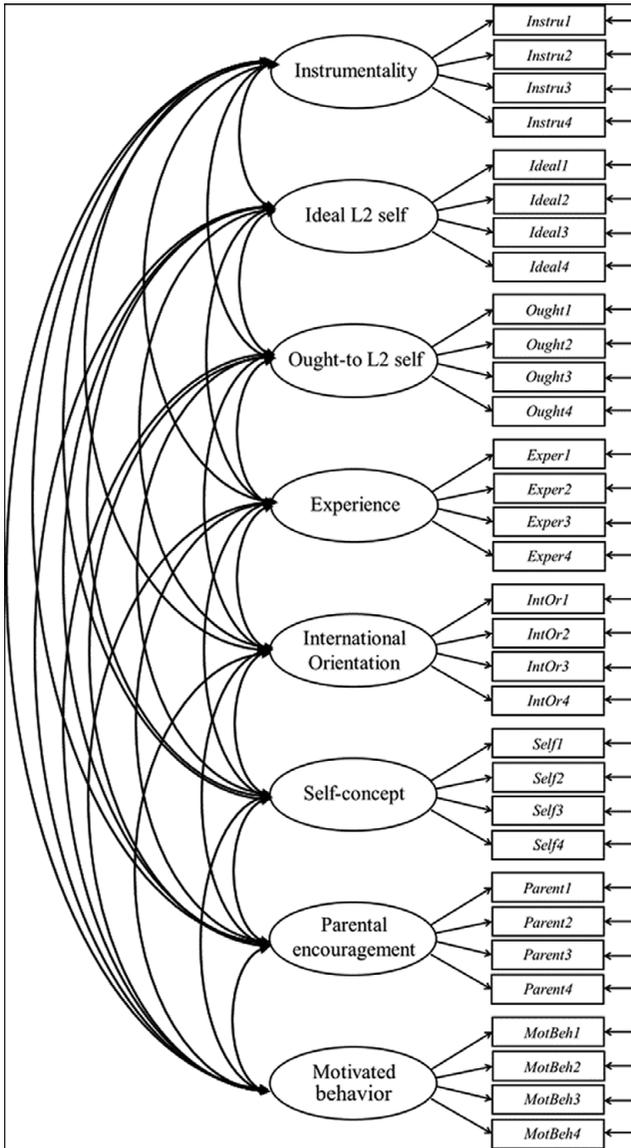


FIGURE A1. Measurement model for motivational traits.

Key to figure A1:

- Circles show the latent variables (motivational scales)
- Boxes show the observed variables (question responses)
- Double-headed arrows represent covariances between each of the latent variables

APPENDIX B**CEFR LEVEL BREAKDOWN BY CLASS FOR EACH SKILL AREAS**

TABLE B1. Listening CEFR levels achieved by class

Class	A0	A1	A2	B1	B2	C	Total
1	0	1	88	80	8	0	177
2	2	0	112	285	33	0	432
3	0	0	47	179	29	3	258
4	0	0	2	156	178	157	493
5	0	0	1	51	127	234	413
Total	2	1	250	751	375	394	1773

TABLE B2. Reading CEFR levels achieved by class

Class	A0	A1	A2	B1	B2	C	Total
1	0	22	119	31	5	0	177
2	0	35	253	130	14	0	432
3	0	11	134	107	5	1	258
4	0	0	20	182	143	148	493
5	0	0	5	88	115	205	413
Total	0	68	531	538	282	354	1773

TABLE B3. Speaking CEFR levels achieved by class

Class	A0	A1	A2	B1	B2	C	Total
1	41	50	60	25	1	0	177
2	33	83	149	165	2	0	432
3	6	26	73	153	0	0	258
4	5	2	26	307	144	9	493
5	1	2	13	186	174	37	413
Total	86	163	321	836	321	46	1773

TABLE B4. Writing CEFR levels achieved by class

Class	A0	A1	A2	B1	B2	C	Total
1	18	80	50	28	1	0	177
2	14	100	151	157	10	0	432
3	4	53	76	118	7	0	258
4	2	1	10	203	238	39	493
5	0	0	6	105	240	62	413
Total	38	234	293	611	496	101	1773

APPENDIX C

SIGNIFICANCE AND T-STATISTICS FOR COMPARISON OF TRAIT SCORES BETWEEN CLASSES

TABLE C1. Factor 1—International orientation class comparisons

	Class1	Class2	Class3	Class4	Class5
Class1		3.98	9.109	5.138	18.655
Class2	P = 0.0001		4.88	0.506	9.093
Class3	P < 0.0001	P < 0.0001		5.284	2.898
Class4	P < 0.0001	P = 0.6128	P < 0.0001		10.428
Class5	P < 0.0001	P < 0.0001	P = 0.0038	P < 0.0001	

TABLE C2. Factor 2—Parental encouragement class comparisons

	Class1	Class2	Class3	Class4	Class5
Class1		2.196	6.598	2.786	10.867
Class2	P = 0.0285		4.784	0.863	6.485
Class3	P < 0.0001	P < 0.0001		3.847	0.643
Class4	P = 0.0055	P = 0.3885	P = 0.0001		4.998
Class5	P < 0.0001	P < 0.0001	P = 0.5202	P < 0.0001	

TABLE C3. Factor 3—Self-concept class comparisons

	Class1	Class2	Class3	Class4	Class5
Class1		5.033	11.199	14.077	33.477
Class2	P < 0.0001		6.724	8.946	21.082
Class3	P < 0.0001	P < 0.0001		0.168	13.385
Class4	P < 0.0001	P < 0.0001	P = 0.8663		15.206
Class5	P < 0.0001	P < 0.0001	P < 0.0001	P < 0.0001	

TABLE C4. Factor 4—Ought-to L2 self class comparisons

	Class1	Class2	Class3	Class4	Class5
Class1		2.463	6.439	2.936	10.296
Class2	P = 0.0141		4.39	0.589	6.035
Class3	P < 0.0001	P < 0.0001		3.872	0.398
Class4	P = 0.0034	P = 0.5562	P = 0.0001		5.206
Class5	P < 0.0001	P < 0.0001	P = 0.6905	P < 0.0001	

TABLE C5. Factor 5—Language learning experience class comparisons

	Class1	Class2	Class3	Class4	Class5
Class1		5.715	14.137	9.965	31.055
Class2	P < 0.0001		7.17	2.118	13.4
Class3	P < 0.0001	P < 0.0001		7.616	4.54
Class4	P < 0.0001	P = 0.0345	P < 0.0001		16.898
Class5	P < 0.0001	P < 0.0001	P < 0.0001	P < 0.0001	

TABLE C6. Factor 6—Motivated behavior class comparisons

	Class1	Class2	Class3	Class4	Class5
Class1		4.195	10.473	7.469	23.082
Class2	P < 0.0001		5.87	0.801	10.479
Class3	P < 0.0001	P < 0.0001		7.168	3.293
Class4	P < 0.0001	P = 0.4232	P < 0.0001		14.345
Class5	P < 0.0001	P < 0.0001	P = 0.0010	P < 0.0001	

TABLE C7. Factor 7—Instrumentality class comparisons

	Class1	Class2	Class3	Class4	Class5
Class1		4.056	10.518	8.857	21.119
Class2	P = 0.0001		6.33	5.059	13.798
Class3	P < 0.0001	P < 0.0001		2.38	9.071
Class4	P < 0.0001	P < 0.0001	P = 0.0176		9.809
Class5	P < 0.0001	P < 0.0001	P < 0.0001	P < 0.0001	

TABLE C8. Factor8—Ideal L2 self class comparisons

	Class1	Class2	Class3	Class4	Class5
Class1		5.451	12.7	13.221	32.397
Class2	P < 0.0001		7.774	6.729	18.582
Class3	P < 0.0001	P < 0.0001		3.614	8.446
Class4	P < 0.0001	P < 0.0001	P = 0.0003		16.5
Class5	P < 0.0001	P < 0.0001	P < 0.0001	P < 0.0001	

APPENDIX D

MPLUS CODE

```
TITLE: Spain - FMM analysis;
DATA: FILE IS Dataset.txt;

VARIABLE:
NAMES ARE                ! names the variables in the dataset
id2 io1 io2 io3 io4 par1 par2 par3 par4
self1 self2 self3 self4 o1 o2 o3 o4
ex1 ex2 ex3 ex4 mb1 mb2 mb3 mb4
instru1 instru2 instru3 instru4
i1 i2 i3 i4 G_V L_scale R_scale
S_scale W_scale;

USEVARIABLES ARE         !specifies which variables to use
id2 io1 io2 io3 io4 par1 par2 par3 par4
self1 self2 self3 self4 o1 o2 o3 o4
ex1 ex2 ex3 ex4 mb1 mb2 mb3 mb4
instru1 instru2 instru3 instru4
i1 i2 i3 i4 G_V L_scale R_scale
S_scale W_scale;

CATEGORICAL ARE         ! indicates categorical observed variables
id2 io1 io2 io3 io4 par1 par2 par3 par4
self1 self2 self3 self4 o1 o2 o3 o4
ex1 ex2 ex3 ex4 mb1 mb2 mb3 mb4
instru1 instru2 instru3 instru4
i1 i2 i3 i4;

IDVARIABLE = id2;      ! declares the id variable

CLASSES = c(5);       ! states the number of latent classes in the
                       ! categorical latent variable

ANALYSIS: TYPE = MIXTURE;
STARTS = 2000 30;
STITERATIONS = 20;
LRTSTARTS = 0 0 2000 30; ! specifies the number of starts and iterations

MODEL:

%OVERALL%              ! sets up relationships between questionnaire
f1 BY io1-io4;         ! responses and the continuous latent variables
f2 BY par1-par4;
f3 BY self1-self4;
f4 BY o1-o4;
f5 BY ex1-ex4;
f6 BY mb1-mb4;
f7 BY instru1-instru4;
f8 BY i1-i4;

OUTPUT:
TECH11 TECH14 TECH10; ! technical output requested

SAVEDATA:
FILE is Filename.txt; ! names output file
SAVE = CPROBABILITIES; ! declares latent class probabilities to be saved
```