

Learners' Profiles for Language Teaching: A Corpus Study of Pakistani Students Errors in Spanish Language

María Isabel Maldonado García
Institute of Languages and Linguistics, University of the Punjab
spanishprofessor1@gmail.com

Abstract

The study measured the effects of building learner profiles for homogeneous groups of students to increase the effectiveness of Spanish language teaching to foreigners. Studies on the effectiveness of learning profiles of L2 students for addressing their language learning needs are almost inexistent. The corpora were collected following an adaptation of the model of Tare et al. (2014). Subsequently, the Tare et al. model for the collection of the corpus was applied. Once the corpus was collected, latest software like OCR Stylus (2022), MAT tagger and AntConc were utilized to analyze the corpus. For this analysis the Ellis error analysis approach (2005) was applied, to reveal common patterns of errors when producing a student output in Spanish language. The results show that Pakistani immigrants present similar characteristics demographically and linguistically and a large number of errors in Spanish language usage derive from their L1. The profile built was tested for a year and utilized for the improvement of teaching and learning materials with positive results. The analysis resulted in a language profile that can be utilized for the teaching of Spanish language to Pakistani learners in any language center by any instructor.

Keywords: Learner's Profile, SLA, Second Language Teaching, Corpus

Introduction

Learner profiles have been utilized for the teaching of various subjects, especially business subjects. While research in the area of Second Language Acquisition (SLA) is vast, learners' profiles have not been used to teach a second language. This research is an endeavor to prove that building learners' profiles for improving instruction and learners' materials for homogeneous groups can assist the teaching of a second language by revealing common patterns of errors, as well as understanding other factors of the group profile such as age, education, mother tongue, etc. which are influencing factors on how students learn a particular language. The researcher has chosen a particular group of students that were learning the Spanish language.

Background

The Spanish Nationality Law of October 15th, 2015, has changed matters for the residents of Pakistani origin in Spain. Due to the law, Pakistani residents who wish to obtain Spanish nationality are now forced by law to achieve a specific level of knowledge of the Spanish language. For this purpose, the Ministry of Interior in collaboration with Instituto Cervantes set up a program for immigrants wishing to receive the nationality so that they can pass a Spanish language proficiency exam called DELE A2 (Maldonado-Garcia, 2018a).

According to Maldonado-Garcia (2018b) literacy success is usually related to a good performance in reading. The Pakistan Economic Survey (2017-2018) states that the literacy rate in the country is 58%. Pakistan's national language is Urdu. Approximately, 8% of the population of Pakistan speaks it as a mother tongue (CIA, 2016) and many more in the country as a second or third language. By a ruling of the Supreme Court (September 2015) it is also the official language of Pakistan. (Maldonado-Garcia, 2015). However, English also has the status of official language and is used as the instruction medium in universities, as well as private schools.

Further, those students in the private schools have a more elevated proficiency level in English since they are taught in this language and they take Urdu separately as a discipline. These individuals are highly fluent in English, may speak Urdu and/or a regional language at home, and obtain better jobs once they have finished their degrees (Maldonado-Garcia, 2017).

Many Pakistanis leave their country looking for better opportunities. Various countries of the world have accepted Pakistani immigrants. The government of Pakistan calls them Overseas Pakistanis. There are communities of Pakistanis in the United States, England, Australia, Oman, Saudi Arabia, and many other countries including Spain. The Ministry of Overseas Pakistanis and Human Resource Development states that approximately amount of 7.6 million Pakistanis live outside of Pakistan (Ministry of OPHRD, 2019). Countries of the Commonwealth will allow them to keep their Pakistani nationality and obtain a second nationality. However, in other countries (mainly Arab countries) they do not have the option of becoming a citizen of that country. In Europe, they would have to give up their Pakistani nationality to attain the nationality of the country where they reside.

Immigrating to a foreign country would consequently entail learning a foreign language (for those of middle and lower classes who do not speak English or those countries where English is not spoken). A study conducted by Maldonado-Garcia (2014) showed that mainly instrumental orientations are at the core of Pakistanis enrolling in foreign language classes. This means that when the average Pakistani national wants to learn a language, there is usually an instrumental motivation behind this desire. These motivations are generally of two kinds: one is to obtain scholarships for study and the second is to obtain employment.

The Pakistani community in Spain mainly resides in Barcelona and according to Beltran Antolín and Sáiz Lopez (2008), the community has more individuals registered in the census than the number of residence permits issued by the government of Spain, which indicates that a part of this community finds itself in an irregular situation.

However, those who are not professionals and have to engage in blue-collar jobs would have to learn English or any other language spoken in the country where they wish to immigrate for survival. The same occurs in Spain. Those Pakistanis who are professionals, in the majority, migrate to places like Canada, USA, and UK (Molina et al., 2015). In these cases, the English language is a facilitating factor. In 2013 it was reported that Pakistanis are the owners of a large number of butcher shops and cell phone shops in La Rambla (Barcelona). The majority of Pakistanis residing in Spain are blue-collar workers who operate stores (grocery, halal

(butcheries), cell phone, etc.) and restaurants. In other cases, they do work for these sorts of businesses.

“As already mentioned earlier, their scarce human capital profile funnels them to the lower jobs of the mainstream labor market: unskilled workers (laborers, factory workers, taxi drivers...), hawkers, or other 'informal' economic activities, sometimes combining several of them. The lack of professional skills of the first generation of Pakistani migrants from Pakistan partly explains why they show a high remarkable tendency toward self-employment and why they have specialized in small businesses.” Molina et al. (2015, p.6)

Literature Review

Originally, learner profiles were used in schools to assess and understand students' needs, strengths and weaknesses to be able to provide a holistic learning environment. Building learner profiles has been utilized in various disciplines, some more extensively than others. Every person belongs to a race, gender, nationality, is a speaker of one or more languages, has an age, goals, etc. These, as well as other factors, make up the elements of an effective learner's profile. Nevertheless, learner profiles have multivariate purposes and can be utilized in many disciplines because they provide benefits for students and teachers in terms of learning success.

In linguistics, and more specifically in second language acquisition, learner profiles have been researched less frequently than in other disciplines. In this case, learner's motivational profiles are more common in SLA, like in Csizer & Dornyei (2005), who studied the motivational profiles of 4,765 Hungarian students according to various factors – such as instrumentality, integrativeness, attitude concerning the speakers of the L2, milieu and linguistic self-confidence, their concern about the culture, and the level of vitality of the community that speaks the L2.

In other cases, the needs were assessed, not only from the perspectives of the students, but also from that of the teacher's as well (Long, 2005; Dias & Dinis, 2014). Needs analysis is a method to build a learner's profile for curriculum design (Brown, 2009); this is another reason and method to engage in learner profiling.

Tare et al. (2014) considered the determination of an individual's interests, educational psychology, needs and goals, proficiency, attitudes, motivation, and language history to build a learner's profile. These factors, in other contexts, have been considered for curriculum design purposes.

In this case, the learner profiles were used for individualizing language training. According to Tare et al. (2013), various characteristics and individual differences play, in fact, a role in learning outcomes. This is the reason why learner profiles are built and utilized.

Concerning learner profiles, it is important to understand that the law of the Official State Bulletin (BOE-Boletín Oficial del Estado) N. 167, Section I, Page 58, 149 which was drafted on the 14th of July, 2015, was to be implemented from the 15th of October, 2015 in Spain and was originally intended towards restoring the nationality of those descendants of the Jewish people who had been expelled from Spain in 1492 to restore relations with this community. Nevertheless, it seems like an unexpected collateral effect, that the Pakistani community is the second most affected community by this law. Garha, Galeano & Valls (2016) believe, the Spanish-Pakistani community in 2014 was 131,230 This is just the number of those who have been recorded. The real number of the residents is not known, as a large number of Pakistanis are illegal and unrecorded in the country's databases. These are irrelevant for this study since only those who are documented can apply for nationality.

Ford et al. (2013) focused their research on highlighting the learning differences among the learners. On the contrary, our study focuses on highlighting the similarities. Some of the characteristics of Pakistani learners seeking the nationality of Spain have been outlined in various studies such as in Maldonado-Garcia (2018a), who determined some of the common features of these learners.

Ijaz et al. (2021) state that many factors affect language and may cause variation. These are race, country of origin, gender, environment and context of non-native use. Although models of building learner's profiles have been written and utilized for curriculum design, as in the studies mentioned earlier, and also for assessing the needs of the learners, an accurate profile of the Pakistani Spanish language learner who seeks nationality of Spain has not been built so far.

It is important, however, to understand the second most important community of immigrants seeking nationality of Spain (the first one is Moroccans), the Pakistani community, in terms of demographics, as well as language needs, goals, and motivation for learning since they are highly misunderstood and under-researched. In this regard, the study investigated their average age, gender, place of residence, motivation types, etc. From a second language acquisition perspective, it is relevant to understand their language history, proficiency levels in the Spanish language for each one of the four skills (speaking, listening, writing, and reading) at the time of the official exam, as well as the kind of errors they commit when engaging in any type of interaction in the Spanish language. Furthermore, an analysis of these errors is necessary for assisting Spanish language professionals when engaging in the instruction of Spanish language to these learners. Hence, this study is an attempt to fill the gap that exists in the current literature by addressing the characteristics of these learners and their language learning needs.

Research Methodology

The importance of the study resides in the fact that the Pakistani origin immigrant is highly under-researched and there is a lack of teaching materials to address their learning needs. The aim was to build a learner profile for teachers of Spanish as an L2 to uncover their demographic as well as language characteristics and be able to address the difficulties this type of learners go through when learning the Spanish language, to ease them, hence, increase their chances of improving their language skills and obtaining nationality. The methodology was based on various steps:

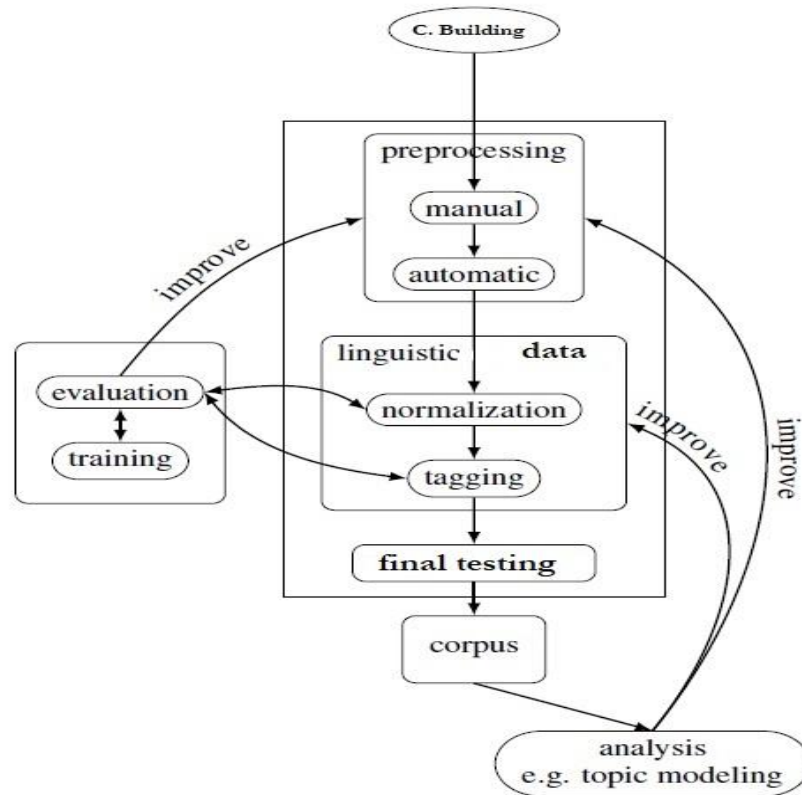
Data Collection

As the research is a corpus-based study, this section provides the steps carried out and the sample techniques followed for the compilation of the corpus. The collection of the corpus was performed through the information the candidates had provided at the time of enrollment. Forty-four enrollment forms were provided by the candidates which contained their gender, age, location, residence, and purpose for doing the exam. This information as well as other demographic information was mentioned in their enrollment forms. In various cases, the candidates sent their enrollment forms by e-mail and also performed their mock examinations online. In the majority of the cases, the candidates visited the office of the examination center

and provided the information on-site. The respondents enrolled in the courses during the years 2015 to 2019 and the forms were collected during this period. After the compilation of the first part of the corpus, data cleaning was performed so that only relevant information remained in the corpus. Inspired by the idea of Cockburn, (2001), we performed a combination and integration of the actual corpus building with data cleaning and analysis, continuously updating the corpus whenever a recurrent problem in data quality was identified. This process revealed that such issues are often detected only during the actual use or compilation of the corpus, so the strategy here was to allow continuous feedback from other stages of processing as well as analysis into corpus building (see Figure 1). Corpus building is divided into three main steps: (a) pre-processing, (b) linguistic annotation of the data, and (c) corpus final testing.

The pre-processing phase included the collection of all three parts of the corpus, into three separate files. The first part, the demographic information of the candidates was to be processed manually so the pre-processing phase only entailed the putting together of all forms of the candidates.

To continue with the second part of the corpus, the error analysis was performed through the mock test the candidates took when they came for enrollment, before taking Spanish language courses in preparation for the official examination DELE A2. The mock tests (pre-test) consisted of linguistic data that targeted the written skills of the respondents. Each test consisted of real-life situations that were posed to the respondents where they had to express themselves by creating a text that fit the requirements of their level. The mock tests consisted of two written exercises of 70-80 words each. Forty-four mock tests were collected.

Figure 1*Process of corpus building*

The respondents enrolled in the courses during the years 2015 to 2019. In this regard, the data was collected during this period. The third part of the corpus was each candidate's A2 DELE result card. Each section of the corpus contained forty-four sets of data that represented the candidates. Each candidate provided a form for their demographic information, a linguistic form for error analysis, and their A2 DELE result card. This was a total of one hundred and thirty-two sets that conformed to the tridimensional corpus. In the pre-processing phase, all three parts of the corpus were compiled, into three different sets of 44 responses. This was done manually for part 1 of the corpus and automatically for parts 2 and 3 of the corpus. The normalization of frequencies was done manually for part 1 of the corpus and automatically for parts 2 and 3. Part 2 was tagged with Multidimensional Analysis Tagger (MAT). The tagger was created by Andrea Nini and is based on the Stanford Tagger for producing an initial Parts of Speech (POS) tagged version of the input. This tagger allowed for the POS usage to be understood in the corpus and frequencies of POS drawn by introducing the resulting tagged file

in AntConc during the process of analysis. The last phase of corpus building was the final testing where the three parts of the corpus were tested for functionality.

Corpus composition:

1. Demographic information: age, gender, place of residence, mother tongue of the respondent, language aptitude, language history, proficiency level in Spanish, and motivation type according to Gardner (1985).
2. Writing sample: Two written texts produced by each candidate, composed of two exercises with clear-cut instructions – they had an approximate number of words of 150 words.
3. A2 DELE result cards: These result cards indicate the results of each oral skill of the candidates – if the candidate passed it is an indication that the learner profile built by the researcher with the purpose of instruction was successful.

Research Instruments

As in Memon et al. (2021) the researcher believes language corpora become useful through the use of computer software that actually processes the corpus files, enabling researchers to obtain the required results in a manner that is easy to understand. The instruments of this research are the software that aided in the analysis. The first instrument utilized to analyze the corpus was the Optical Character Reader, (OCR) Stylus (2022) software that was utilized to convert handwritten text into digital text to be able to analyze it with another software. The second one was the Multidimensional Analysis Tagger (MAT) that was used to tag the corpus for a later analysis. Finally, AntConc was used to separate parts of speech into analyzable, independent units and find out frequencies.

Ethical Considerations

International ethical considerations for human participants were followed (Belmont Report' of 1979, from the US FDA's National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research that enlisted the ethical guidelines for human research subjects protection). Further, the research is independent and impartial and was not funded by any of the concerned organizations and institutions.

Limitations of the Study

The research is limited to building an average second language learner profile of a Pakistani resident of Spain who applies for Spanish nationality. For this reason, only A2-level Pakistani language learners, who sought nationality, were selected for the study, and the profile was built according to the relevant information and the criteria set for this purpose.

Steps of the Methodology

To build a learner profile various criteria were explored. This was done through a process of triangulation by utilizing an experimental mixed methods approach. The theoretical framework includes, in the first step, the collection of demographic information for each candidate, such as in Tare et al. (2014) – although, Tare et al.’s model was adapted to fit the needs of this study. In the second step, a corpus was built according to the corpus-building steps mentioned above.

In the second step, a pre-test was conducted, where a sample of the respondents’ writing was collected, which was made into a corpus for analysis. This corpus was their pre-test. The approach through which this data was analyzed were the Ellis models (1994/2005) as well as (Maldonado-Garcia & Haider, 2023) which are an adaptation of Corder’s error analysis model (1981). The model includes five elements. These are 1) Sample collection, where the definition and sample of the study is decided; 2) Identification phase, where the errors are identified. This was done through the use of the Stylus software (2022) which is a text analyzer; 3) Description of the errors, where errors are described and the type of errors are mentioned. This step was performed manually; 4) Explanation of the errors, where the errors are explained and classified according to grammar; and 5) Evaluation, where the reasons for the errors are explained.

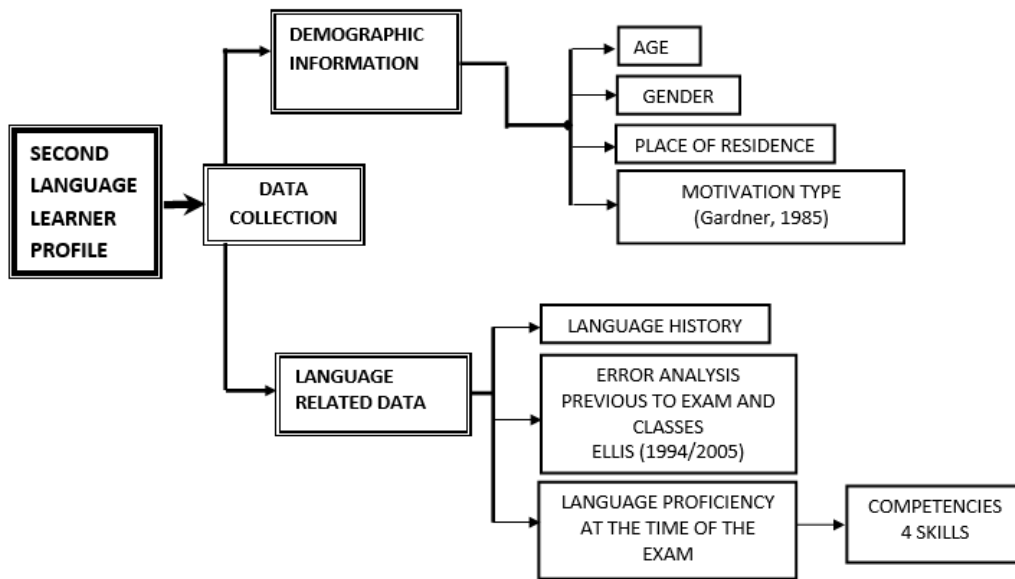
The post-test was not a mock test like earlier, but was their original DELE A2 examination result. A written exercise with the same format was administered during the real-life exam. In this step, a thorough analysis of their errors was not made after instruction; rather, their ability to pass after instruction was considered, which meant the number of errors had been reduced and the learner’s profile was useful.

An analysis of the result cards of the DELE A2 learners after instruction – also through their mock exams (not their DELE A2 exams), and other information collected at the time of enrollment for preparatory courses – were used to build the language profile. The addition of extra elements to the learner profile is done to create a more comprehensive profile than those previously explored for other types of learner profiles. It is also relevant to investigate which errors are the most common in this type of learner and why those errors are committed.

It is believed that in the majority of the cases, certain errors have their origin in the interference with the learners' L1 – in this case, Urdu and Punjabi mostly. The following diagram (Fig 2) illustrates the methodology:

Figure 2

Methodology for Building a Learner’s Profile for Second Language Learners



In addition, once the learner profile was made, it was tested through a survey of 5 Spanish language teachers who applied the model to their teachings during the year 2020.

Analysis of the Data, Results and Discussion

As per the methodology steps mentioned before, the researcher proceeded with the data collection.

Data Collection

The data collection for Table 1 was performed by utilizing the registration forms that the respondents had filled out themselves at the time of enrollment for the preparatory courses for the DELE A2 exams. In the first step towards building a Second Language Learner Profile, the gender, age, city of permanent residence, and motivation for learning Spanish of the respondents were investigated. Gender, age, and city of permanent residence are different factors that affect the Spanish language acquisition of the learners, as well as their motivation for learning the language.

It is well known that age is a factor in the learning of a second language. Talking about gender, for the Pakistani community, women staying at home is the normality, and hence, have less interaction with Spanish people, making their language immersion limited or somehow impaired, while in the case of the males since they engage in outside employment it may or not be limited, depending on the factors that condition their employment. These can be working in an environment that facilitates interaction with Spanish speakers or, on the contrary, working in a location where they have access only to other Pakistanis, hence with limited or inexistent interaction. The place of residence also conditions the learning of the speakers. It is not the same to live surrounded by Spanish language speakers as to live in a city or village in Pakistan where no one speaks the Spanish language.

Results and Discussion

The results that were drawn from the first part of the corpus are as follows:

Table 1

Demographic Information

Respondent #	Gender	Age (At the Time of Enrollment)	Country of Residence	City of Residence	Motivation For Learning Spanish Language
1.	M	40	Spain	Barcelona	Obtaining nationality
2.	M	48	Spain	Barcelona	Obtaining nationality
3.	M	39	Spain	Barcelona	Obtaining nationality

4.	M	36	Spain	Barcelona	Obtaining nationality
5.	M	47	Spain	Barcelona	Obtaining nationality
6.	M	43	Pakistan	Sialkot	Obtaining nationality
7.	M	35	Spain	Barcelona	Obtaining nationality
8.	M	37	Spain	Barcelona	Obtaining nationality
9.	M	52	Spain	La Coruña	Obtaining nationality
10.	M	45	Spain	Barcelona	Obtaining nationality
11.	M	31	Spain	Barcelona	Obtaining nationality
12.	M	41	Pakistan	Multan	Obtaining nationality
13.	M	32	Pakistan	Sahiwal	Obtaining nationality
14.	M	33	Spain	Barcelona	Obtaining nationality
15.	M	25	Spain	Barcelona	Obtaining nationality
16.	M	38	Spain	Barcelona	Obtaining nationality
17.	M	37	Spain	Barcelona	Obtaining nationality
18.	M	43	Spain	Barcelona	Obtaining nationality
19.	M	39	Spain	Barcelona	Obtaining nationality
20.	M	48	Spain	Barcelona	Obtaining nationality
21.	M	41	Spain	Barcelona	Obtaining nationality
22.	F	37	Pakistan	Gujranwala	Obtaining nationality
23.	M	44	Spain	Barcelona	Obtaining nationality
24.	M	49	Spain	Barcelona	Obtaining nationality
25.	M	27	Spain	Barcelona	Obtaining nationality
26.	M	21	Spain	Barcelona	Obtaining nationality
27.	M	43	Spain	Barcelona	Obtaining nationality
28.	M	49	Spain	Barcelona	Obtaining nationality
29.	M	45	Spain	Barcelona	Obtaining nationality
30.	M	36	Spain	Barcelona	Obtaining nationality
31.	M	46	Spain	Barcelona	Obtaining nationality
32.	M	25	Spain	Barcelona	Obtaining nationality
33.	M	50	Spain	Barcelona	Obtaining nationality
34.	F	42	Spain	Barcelona	Obtaining nationality
35.	M	42	Spain	Barcelona	Obtaining nationality
36.	M	37	Spain	Barcelona	Obtaining nationality
37.	M	43	Spain	Malaga	Obtaining nationality
38.	M	36	Spain	Barcelona	Obtaining nationality
39.	M	40	Spain	Barcelona	Obtaining nationality
40.	M	37	Spain	Barcelona	Obtaining nationality
41.	M	39	Spain	Barcelona	Obtaining nationality
42.	F	32	Spain	Madrid	Obtaining nationality
43.	M	47	Spain	Barcelona	Obtaining nationality
44.	M	46	Spain	Barcelona	Obtaining nationality

The previous table displays various aspects that have been collected separately as well in the following tables:

Table 1.2

Gender Frequency

Gender	Frequency	Percentage
F	3	6.9
M	41	93.1
Total	44	100.0

The results show that out of 44 learners 41 were male and 3 were female. It can be understood that 93.1% of the learners (nationality seekers) are male and only 6.9% of the learners are female.

Table 1.3

Age Frequency

Age	Frequency	Percent
21	1	2.3
25	2	4.5
27	1	2.3
31	1	2.3
32	2	4.5
33	1	2.3
35	1	2.3
36	3	6.8
37	5	11.4
38	1	2.3
39	3	6.8
40	2	4.5
41	2	4.5
42	2	4.5
43	4	9.1
44	1	2.3
45	2	4.5
46	2	4.5
47	2	4.5
48	2	4.5
49	2	4.5
50	1	2.3
52	1	2.3
Total	44	100.0

The table shows that the great majority of the respondents are in the age bracket of 36 to 49. It can be safely stated that a large percentage of the learners are middle aged.

Table 1.4

Country of Residence

Country of Residence	Frequency	Percent
Pakistan	4	9.1
Spain	40	90.9
Total	44	100.0

This table shows that 90.9% of the learners have permanent residence in Spain. This was expected of nationality applicants.

Table 1.5*City of Residence*

City of Residence	Frequency	Percent
Barcelona	37	84.1
La Coruña	1	2.3
Madrid	1	2.3
Málaga	1	2.3
Gujranwala	1	2.3
Multan	1	2.3
Sahiwal	1	2.3
Sialkot	1	2.3
Total	44	100.0

The above information indicates that 84.1% of the learners have their residence in Barcelona, while a minority (6.9%) lives in other cities of Spain and the same number (6.9%) reside in cities of Pakistan (with legal residence in Spain). This result was clear as Pakistani residents are mainly based in selected Barcelona neighborhoods.

Table 2*Language Related Data*

Respondent #	Mother Tongue	Additional Languages Spoken (Other than Spanish)	Prior Spanish Classes Taken
1.	Punjabi	Urdu, Some English	No
2.	Punjabi	Urdu	No
3.	Punjabi	Urdu	No
4.	Punjabi	Urdu,	No
5.	Punjabi	Some Urdu	No
6.	Punjabi	Urdu	No
7.	Punjabi	Urdu	No
8.	Punjabi	Urdu	No
9.	Hindco	Urdu	No
10.	Hindko	Punjabi, Urdu, and English	No
11.	Hindko	Urdu, English	No
12.	Siraiki	Punjabi, Urdu	No
13.	Punjabi	Urdu, English	No
14.	Punjabi	Urdu, German, and some English	No
15.	Punjabi	English, Urdu	No
16.	Punjabi	No	No
17.	Punjabi	Urdu	No
18.	Hindko, Punjabi	Urdu	No
19.	Punjabi	Urdu	No
20.	Punjabi	Urdu	No
21.	Pushtu	Some Urdu	No
22.	Punjabi	English, Urdu	No
23.	Punjabi	Urdu	No
24.	Hindco	Urdu,	No
25.	Punjabi	Urdu, English	No
26.	Urdu	Pashto	No
27.	Hindko, Punjabi	Urdu	No

28.	Urdu	English	No
29.	Punjabi	Urdu	No
30.	Punjabi	Urdu, English	No
31.	Hindco	Urdu, English	No
32.	Punjabi	Urdu, English	No
33.	Punjabi	Urdu, English.	No
34.	Punjabi	Urdu, English	No
35.	Punjabi	No	No
36.	Punjabi	Urdu	No
37.	Punjabi	English, Urdu	No
38.	Punjabi	Urdu	No
39.	Punjabi	Urdu	No
40.	Punjabi	Urdu	No
41.	Hindco	Urdu	No
42.	Punjabi	Urdu, English	No
43.	Punjabi	Urdu, Some English	No
44.	Punjabi	Urdu, Some English	No

Table 2 demonstrates that the respondents are mainly Punjabi (L1) speakers who also have learned other languages such as Urdu and some of them have learned English. There is a minority who speaks Hindco as an L1 and others who speak Urdu, Pashto, and Siraiki. The great majority speak Urdu as an L2 and some of them also speak English to some extent. In addition, none of the learners had taken prior classes in the Spanish language. Until now, all data had been drawn manually by the researcher from the forms that constituted the first part of the corpus. At this point, in the pre-test, the researcher proceeded to the identification of errors that were performed through the mock exams taken at the time of enrollment (different times for each learner) which were compiled into the second part of the corpus.

The exam consisted of a written assignment of fifty minutes duration with two written exercises. Each written exercise consisted of 70 to 80 words, with a maximum of 160 words per mock exam. This part of the corpus consisted of approximately 7,040 words. The format was the same as that of the written part of the official exam. The identification of errors was done through the conversion of hand-written texts to digitally written texts through the use of OCR (Optical Character Reader) Stylus technology. This application enabled the researcher to scan the second part of the corpus and convert it to editable texts with the click of a button. Once this process was completed, Stylus (2022) was utilized – a software to detect grammatical and orthographical errors in Spanish. Stylus is an application that performs a tridimensional output. It uses a verb conjugator for the analysis of verbs, a morphosyntactic analyzer, and a reverse dictionary. The conjunction of these three analyses ensures the detection of any type of error in a Spanish

language text. The use of the Stylus application enabled the grammar, punctuation, and orthography errors to be identified and highlighted, to be mapped on tables later on. AntConc was used to find the frequencies of each part of speech and through this the frequency of errors, which was part of the error analysis process. Table 3 depicts the errors each respondent presented.

Table 3

Error analysis. Identification of errors from the corpus

Respondent #	Identification of Errors
1.	Sentence structure, spellings, tenses, lack of articles.
2.	Absence of articles, gender assignment, spellings, verbs (tense)
3.	Articles, verbs (tense), gender disagreement, spellings, punctuation
4.	Lack of articles, tense errors, spelling, word order
5.	Absence of articles, spelling errors, word order, verbs (tense), gender assignment
6.	Person-tense assignment, spelling errors, plural-singular inconsistencies, verb errors, lack of articles
7.	Lack of articles, tense confusion, por/para confusion, lack of prepositions, gender agreement
8.	Lack of articles
9.	Spelling errors, tense errors
10.	Punctuation, lack of articles, tense errors, spelling
11.	Lack of articles, punctuation errors, lack of plurals.
12.	Sentence incoherence, lack of articles, prepositions, and conjunctions, verbs (tense)
13.	Punctuation errors, lack of some articles, tense and gender assignment errors.
14.	Spelling errors, plurals, lack of prepositions, lack of articles,
15.	Lack of articles.
16.	Verbs (Confusion ser/estar), absence of articles, spelling errors,
17.	Spelling errors, tenses errors, lack of articles, lack of prepositions
18.	Spelling errors, gender assignment confusion, tenses errors, sentence incoherence, lack of prepositions
19.	Spelling errors, word order, person-tense relation, sentence incoherence
20.	Spelling errors, lack of articles, plurals.
21.	Lack of articles, tenses, gender assignment, spellings, plurals, lack of preposition
22.	Lack of articles and prepositions, spelling errors, tense errors, punctuation errors.
23.	Spelling errors, tense errors, articles
24.	Coherence problems. Unable to write a coherence sentence
25.	Spellings, lack of prepositions,
26.	Spellings, sentence incoherence,
27.	Spelling errors, gender assignment confusion, tenses errors, sentence incoherence, lack of prepositions
28.	Word order issues, spelling errors,
29.	Switch of diphthong vowels, spelling errors.
30.	Spelling errors, tense and person error, lack of articles
31.	Spelling errors, lack of prepositions,
32.	Word order issues, tenses errors, sentence incoherence
33.	Spellings, sentence incoherence, lack of articles, tenses errors,
34.	Spellings, tenses errors, lack of articles, preposition errors,
35.	Lack of articles, spelling errors, verbs errors (tener for estar) Comprehension problems
36.	Lack of articles, tenses (ser/estar confusion) reflexive particles misuse
37.	Tenses errors, spellings, lack of prepositions, lack of articles
38.	Coherence problems. Unable to write a full sentence.

39.	Structur errors, spelling errors, lack of articles, conjunctions and prepositions. Tense errors.
40.	Structural errors, lack of prepositions, articles, and some tenses.
41.	Spelling, gender assignment,
42.	Spelling errors, lack of articles, plurals.
43.	Spellings, lack of articles, tenses, word order
44.	Spellings, word order, lack of articles, tenses

After the type of errors were identified, the description of the errors followed. The description was also performed using the Stylus morphosyntactic analyzer. After that, the analysis continued with the explanation and the evaluation phases of the analysis which was performed manually by the researcher:

Table 4

Ellis' model application. Error analysis

Error #	Collection	Identification Description	+	Explanation	Evaluation
1	33 Candidates	Spellings		Errors in the formal writing of a word. The candidate tries to write how they hear applying the rules of their mother tongue.	This error occurs because the candidates have not received formal education in the L2/TL. For this reason, since they have orally learned the language by listening and have not developed their written proficiency, this error is present.
2	7	Sentence incoherence		This error is usually a collection of errors (which have also been explained separately). Usually, lack of articles + incorrect word order + lack of prepositions.	The lack of articles is an interference of Urdu/Punjabi since these languages do not present articles and hence, the student cannot understand the usage of articles in the L2. Word order in Spanish presents an SVO structure, while in Urdu/Punjabi an SOV structure, hence the students place the object before the verb in Spanish creating an erroneous structure. The absence of prepositions is also an interference from the L1. Prepositions are inexistent in Urdu/Punjabi (Maldonado-Garcia, 2013) but these languages show the presence of postpositions. Hence, this is as well a negative transfer from the L1.
3	30	Lack of articles		The candidate places nouns without definite or indefinite articles, creating	The lack of articles is an interference with the L1. Urdu/Punjabi does not present definite articles. Indefinite

			a void in the structure of the sentence.	articles are usually represented through the numeral "aik" since in Urdu/Punjabi there are no indefinite articles either.
4	6	Plural omissions and agreement	The observed plural omissions are of two kinds: 1. Omission in the –s/-es at the end of the noun/adjective. 2. Article-noun/adjective agreement, usually making a construction where the article is singular and the noun/adjective plural.	The plurals in the Spanish language are represented by adding –s and –es at the end of words. Urdu/Punjabi add a different vowel to represent plurals. Since Urdu/Punjabi do not have articles this is a common error. Also, students usually do not understand there has to be an agreement between all the elements of the plural construction article+noun+adjective.
5	28	Tenses errors + erroneous verb application	The error manifests in the application of the wrong form of the verb or an erroneous verb composition by the learner.	There is a variety of reasons why this error happens. Mainly because the mother tongue of the respondents doesn't have a variety of past tenses and there is no correspondence as such in the Spanish language. It also happens because in Urdu/Punjabi age and other compositions are made with a different verb.
		Person-tense assignment	The error manifests when the tense used by the learner belongs to another grammatical person. For example: Yo comiste. This is = to 1 st person + verb in 2 nd person.	This error takes place because in Urdu/Punjabi while there are grammatical persons, they are not all different from each other, while in Spanish they are all different.
		Ser/estar confusion	The error manifests by the arbitrary use of ser/estar without following the proper rules. For example; "Él <u>es</u> en la estación" instead of "Él <u>está</u> en la estación." The verb "estar" is used in Spanish to denote location.	The Spanish language has two forms of the verb to be. The great majority of learners who do not have a Romance language as their mother tongue will display this error.
			The error manifests by the use of ser when stating age. In Spanish age is expressed with the verb to have. For example: "Mi	This error is an influence of the L1 as well. This expression with "tener" is made in Urdu/Punjabi with "ser."

		Incorrect use of ser+age construction.	hijo es cinco años” instead of “Mi hijo tiene cinco años” which is the correct expression.	
6	7	Word Order	The error manifests as adjectives before nouns and object before verbs making the sentence structure confusing	In Spanish language adjectives are placed after the noun while in Urdu/Punjabi the adjectives go before the noun. SVO vs SOV.
7	14	Lack of prepositions	No prepositions mean a void in meaning in some sentences.	Respondents display confusion in the usage of prepositions because Urdu/Punjabi do not have prepositions, but rather postpositions. This makes it difficult for the learner to adjust to the use of prepositions.
8	9	Grammatical gender assignment	The error manifests by a change in vowel at the end of the noun or adjective, or a gender disagreement in the construction of article+noun adjective genders.	Gender assignment is different in Urdu/Punjabi and Spanish. For this reason, since both languages use an arbitrary gender assignment system, problems arise.
9	5	Punctuation	Lack of commas, full stops, capital letters, colons, and semi-colons.	Many of the learners do not have an understanding of punctuation rules. Since they have not been taught the language formally, and usually they have not reached a high level of education, they are unable to understand the use of commas, capital letters, colons, and semi-colons. Full stops are easier for them. This is also an interference from their L1 which lacks punctuation signs.
10	1	Switch of diphthongs vowels	The error manifests by the switch in the order of the vowels. For example: “sies” instead of “seis.”	This error has to do with phonetics. The learner writes what they think they hear which in Spanish is not written as they perceive.
11	1	Por/para confusion	The error is noted when the wrong preposition is utilized.	Since both prepositions are translated as “for” the learner is unable to distinguish when to use one or the other.

The errors can be classified into two different types: inter-lingual errors which are based on negative transfer from L1 or interference from their L1 (either Urdu or Punjabi in this particular case) and intra-lingual errors that are within language factors, also known as

developmental errors, and that represent lack of understanding of the L2 grammar rules or their overgeneralization.

The next table depicts the exam scores of the respondents, in other words, the results of the post-test. These scores were obtained after the candidates had received instruction at the institute for a minimum of three months. In this step, the third part of the corpus is involved. The DELE A2 result cards were compiled into a text file that had 44 pages, one page per respondent. Through a frequency analysis of the words representing each language learning skill – reading, writing, listening, and speaking – the scores of each individual were compiled into columns and mapped in Table 5. In the pre-test, the written samples were collected and the errors were highlighted. It was then clear that most of the candidates would have failed the official written test, given the frequency of the errors found. After a minimum of three months of instruction (in some cases four) the respondents presented the following scores which are a reflection of how proficient each candidate was in each language skill:

Table 5

Scores of Respondents

Respondent Number	Reading	Writing	Listening	Oral Skills	Passed/Failed
Max. Marks	25	25	25	25	P/F
1.	16.67	14.73	11.67	19.17	P
2.	17.5	18.06	20.0	24.38	P
3.	15.0	17.43	15.0	18.54	P
4.	11.67	20.28	10.83	24.38	P
5.	17.5	15.34	20.0	17.92	P
6.	2.5	13.55	13.33	17.29	F
7.	15.0	16.67	20.0	23.75	P
8.	15.0	15.0	15.0	15.0	P
9.	16.67	15.49	18.33	18.54	P
10.	16.67	14.17	17.5	20.42	P
11.	18.33	16.67	25.0	23.75	P
12.	19.17	15.91	20.0	23.75	P
13.	8.33	13.82	6.67	23.75	F
14.	15.83	17.92	17.5	25	P
15.	24.17	25.0	25.0	24.38	P
16.	12.5	0.0	18.33	23.75	F
17.	6.67	5.63	9.17	22.5	F
18.	16.67	11.94	14.17	17.92	F
19.	21.67	13.27	25.0	17.29	P
20.	13.33	17.08	17.50	24.38	P
21.	12.5	17.5	10.83	23.75	P
22.	13.33	10.35	15	17.92	F
23.	10.00	7.11	5.00	22.50	F
24.	12.50	5.15	7.50	17.92	F
25.	20.0	17.01	22.50	24.38	P

26.	8.33	8.96	8.33	17.29	F
27.	7.50	5.56	15.83	21.13	F
28.	17.50	14.58	14.17	21.88	P
29.	24.17	19.44	20.83	24.38	P
30.	14.17	16.67	10.83	19.17	P
31.	19.17	22.01	11.67	24.38	P
32.	20.00	20.90	15.00	19.17	P
33.	14.17	20.01	9.17	20.83	P
34.	16.67	13.89	8.33	17.92	F
35.	11.67	18.33	12.50	23.13	P
36.	22.50	10.35	20.00	19.17	P
37.	15.00	17.22	16.67	22.50	P
38.	10.83	9.51	14.17	20.42	F
39.	6.67	9.32	14.17	23.75	F
40.	14.17	16.67	16.67	23.75	P
41.	6.67	3.77	10.83	15.42	F
42.	8.33	21.67	10.00	23.75	P
43.	7.50	14.45	14.17	19.17	F
44.	10.00	8.89b	14.17	17.92	F

Legend: Candidates who passed: 28 Candidates with higher scores in productive skills: 32 Candidates who failed: 16

To pass one skill, the candidate has to achieve a minimum average mark of 12.5. However, reading is graded jointly with writing and if the candidate achieves more than 30 marks in those two sections, the candidate has passed the reading-writing sections, according to the Instituto Cervantes rules. The minimum mark by joining both parts is 60 – 30 for each two-skill section. Speaking and listening are grouped and the same process applies. In the above table, the writing skills marks which is the skill of the mock test or pre-test. The production skills of any candidate are writing and speaking; this means that the candidate has to produce text in the case of writing and speech in the case of speaking. The writing skill is also an indication of the speech a candidate can produce in terms of grammar and correctness. These skills are related because they are productive. The marks of the oral skills have been highlighted as well. Reading and listening are reactive skills. This means that the candidate is given text and speech and he has to return a set of answers according to what he/she read or heard. Instruction was given at Al-Andalus Institute for all skills. However, the pre-test had been taken on the writing skill only since writing is a productive skill and an indicator of the speech the candidate can produce. The table indicates average scores in the productive skills that are higher than those of the receptive skills. The pre-test indicates that initially the respondents had a significant number of errors and made many grammar and orthography mistakes, while during the post-test, these scores showed that 32 candidates out of 44 had achieved higher than average scores in their writing skills (which was reflected also in their oral skills) where all candidates had achieved a high score and they all passed in these productive skills. This is also an indication that the language profile was useful

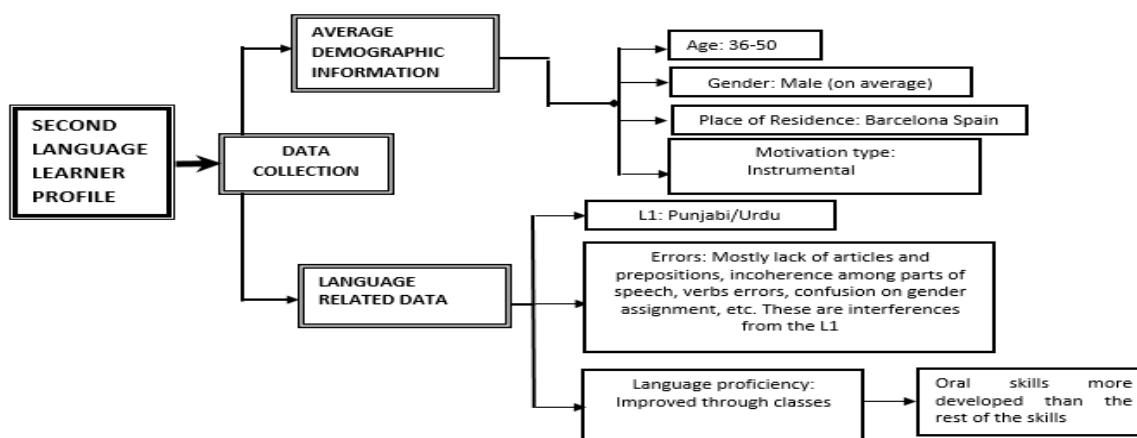
during instruction, as the profile mostly targeted productive skills. The table indicates that twenty-eight candidates passed the exam. The post-test shows that 72% of the candidates had increased scores after instruction.

Profile of the Pakistani nationality seeker of Spain: Discussion.

The results of this study put on display various facts. The Pakistani nationality applicant of Spain is on average male and middle-aged. He lives in Barcelona the majority of the time and the errors he makes when communicating in Spanish are mostly related to interferences from his L1. His speaking skills are more developed than the rest of his language skills (writing, reading, and listening); this is so since usually this learner is engaged in some type of work or employment where he never engages in writing (which is the least developed of all his skills) and reading and there is a limited amount of listening. On the other hand, his job may require him to mostly engage in speaking. This type of learner has not taken Spanish language classes before and if it weren't because of the exam, he would probably never take them. Learning Spanish is an obligation for him, the means to achieve his goal of receiving nationality. For this reason, it is an instrumental motivation that drives him to take the necessary classes, as proven in earlier research (2014). It could be thought that the motivation could be integrative; however, the results prove that the only purpose of this learner is to receive the Spanish nationality, not to learn the language and not to integrate into the Spanish community, as he has a community of his own where the majority of the times engagement occurs in his native language – either Urdu or mostly Punjabi. The following figure (Fig 3) illustrates the obtained (average) learner's profile:

Figure 3

Results. Learner Profile: The Pakistani Nationality Applicants of Spain



Testing the model

After the learner profile was made and used for instruction at the Al-Andalus Institute of Languages, Lahore, it was distributed among the teachers of the Spanish language for testing.

A small survey of 5 teachers of the Spanish language was taken where 100% of the respondents stated that the learner profile had assisted them in preparing appropriate curriculum and teaching materials, geared towards addressing the needs of those students.

According to the teachers, all segments of the profile built were useful. Furthermore, the teachers reported that the analysis of errors was most useful, as they could understand ahead of time the particular language problems of this group of students. Because these learners are in their majority male and middle-aged, the instructors put together age-appropriate topics and materials that addressed the errors, they knew learners would make according to their profiles.

Keeping in mind that in Urdu and Punjabi languages, the definite article does not exist; therefore, special lessons were developed to address this issue. Further, verb lessons were given to the students for the differentiation of "ser" and "estar", as well as the different forms of the past tense and conditional. The issues of word order and sentence incoherence were addressed by explaining the Spanish language word order when the problem took place. Another common error was gender assignment. Gender assignment in Spanish language is a grammar element that derives from Latin without any particular rule in the assignment. It is easy to recognize from the a/o endings of words, however, not in all cases. The rule here is to look at the article to find out the gender of a particular word. In this case, relevant exercises were provided to the learners where they were able to address their gender assignment learning issues.

Since not all learners had the same frequency of errors, particular importance was given to those errors that derived from their L1 (namely Urdu and Punjabi). In addition, the fact that the teachers were able to speak an intermediate level of Urdu and understand Punjabi, this assisted in the understanding of the errors committed by them.

Further, a considerable improvement was observed in the performance of those who had failed the test in previous examinations. This proved the effectiveness of the learner's profile and brought necessary improvement to the teaching and learning of the Spanish language of this group.

After a year of implementation, the teachers reported the profile was useful in addressing the language problem of Pakistani immigrants making it easier for them to learn the Spanish language in regards to demographics as well as other relevant aspects at the time of enrollment and after the A2 DELE test.

Conclusion

Individualizing instruction is an important factor in receiving effective and personalized teaching (Connor et al., 2009). Furthermore, learning how to make language learner profiles is fundamental for instructors of foreign languages to better understand and address the needs of the learners, and design curricula according to these needs.

A language profile of the Pakistani nationality applicant was built to understand the language needs of this community of learners. The profile revealed a set of patterns that are similar in the language learners of this particular background. It was proven that the learner profile assisted educators in building appropriate and relevant curriculums and creating teaching materials in Spanish as a second language for this type of learner. The learner profile resulted in teachers who can better understand the needs of this type of learners and a better learning experience and progress of the students. The learner profile built in this research was tested for one year through the design of materials directed to address the errors revealed. It proved useful and efficient for those teachers who used it.

References

- Cockburn, A. (2001). *Agile Software Development*. Addison-Wesley Professional, Boston, USA.
- Connor, C. M., Piasta, S. B., Glasney, S., Schatschneider, C., Crow, E., Underwood, P., Fishman, B., Morrison, F. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development*, 80, 77-100.
- Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Csizer, K., Dornyei, Z. (2005). Language Learners' Motivational Profiles and Their Motivated Learning Behavior. *Language Learning*, 55(4); 613–659.
- Dias, S. B., & Diniz, J. A. (2014). Towards an enhanced learning management system for blended learning in higher education incorporating distinct learners' profiles. *Journal of Educational Technology & Society*, 17(1), 307-319.
- Ellis, R. & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford: Oxford University Press.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ford, K. L., Cabell, S. Q., Konold, T. R., Invernizzi, M., & Gartland, L. B. (2013). Diversity among Spanish-speaking English language learners: Profiles of early literacy skills in kindergarten. *Reading and Writing*, 26(6), 889-912.
- Gardner, Robert C. (1985) *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London: Arnold.
- Garha, N. S., Galeano, J., & Valls, A. D. (2016). South Asian immigration to Spain: Socio-demographic profile and territorial distribution, 2000-2014. *Asian and Pacific Migration Journal*, 25(2), 191-205. <https://doi.org/10.1177/0117196816639166>
- Government of Pakistan (2018) *Overseas Pakistanis and Human Resource Development Report*. Islamabad: Ministry of OPHRD. <https://www.ophrd.gov.pk/Index.html>

Government of Pakistan (2019) *Pakistan Economic Survey*. Islamabad: Ministry of Finance.
http://www.finance.gov.pk/survey_1718.html

Ijaz, M., Tahir, A., & Ahmed, S. (2021). A Corpus-based Multidimensional Analysis of Linguistic Variation in Pakistani Newspapers Columns of Opinion during Covid-19. *Corporum: Journal of Corpus Linguistics*, 4(2), 59-75.

Long, M.H. (2005). *Second language needs analysis*. Cambridge: Cambridge University Press.

Maldonado-García, M. I. (2013) *Comparación del Léxico Básico del Español, el Inglés y el Urdu*. Doctoral Thesis. Madrid: UNED. 183.

Maldonado-García, M. I. (2014a). The Urdu Language Reforms. *Almas*, 15, 14-24

Maldonado-García, M. I. (2014b). Foreign Language Learning Orientations. The Case of Pakistan. *Annual Pakistan Journal*. 50, 1-14. https://www.pscpesh.org/PDFs/PJ/Volume_50/01_Garcia.pdf

Maldonado-Garcia, M.I. (2015) Reforming the National Curriculum. Teaching the National and Provincial Languages of Pakistan. *Journal of Elementary Education*, 25(2) 73-87.
[http://pu.edu.pk/images/journal/JEE/PDF-Files/5_Maria%2025\(II\).pdf](http://pu.edu.pk/images/journal/JEE/PDF-Files/5_Maria%2025(II).pdf)

Maldonado-Garcia, M.I. (2018) Debate on Urdu as the Official Language of Pakistan: English versus Urdu. *Almas*, 18, 25-38. <https://scholar.google.com/scholar?oi=bibs&cluster=16775287100653216561&btnI=1&hl=en&authuser=1>

Maldonado-Garcia, M.I. (2018a) Politics of Immigration And Language: The Case of Pakistani Residents In Spain. *International Journal of English Linguistics*, 8 (3) 36-46. <https://ccsenet.org/journal/index.php/ijel/article/view/73375>

Maldonado-Garcia, M.I. (2018b) Improving University Students' Writing Skills in Pakistan. *The European Educational Researcher*, Vol 1. (1) 1-16. <https://eu-er.com/download/improving-university-students-writing-skills-in-pakistan-11592.pdf>

- Maldonado-Garcia, M.I., & Haider, M. (2023). The Role of Diacritics in Reading Urdu. Can children read without “the dots”? *Linguistic Forum: A Journal of Linguistics*, 5(2), 22-31.
- Memon, M. A., Pathan, H., & Memon, S. A. (2021). An Intercultural Investigation of Interactive Metadis course Markers in Research Articles by Pakistani & British Engineers. *Corporum: Journal of Corpus Linguistics*, 3(2), 51-72.
- Molina, J. L., Valenzuela, H., García-Macías, A., Pamplona, J., & Lubbers, M. (2015). Measuring Social Capital in Ethnic Enclaves: Indians in Lloret de Mar, and Pakistani in Barcelona city. *Handbook of Research Methods and Applications on Social Capital*. Edward Elgar Publishers. (Pre-print Edition).
- Tare, M. et al. (2014). *Building a language learner’s profile. Characteristics which inform training and pedagogy*. USA: University of Maryland.