



BRITISH
COUNCIL

A :: FNOAG//XCKPW?9BE//17B=6>?93EB17B=ZV28TZQU-2535%AXG
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
83QJHY4HS83Q%AXGX981^9#98>T//GJB::83QJHY4HS83Q%AXGX98
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
A::FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%AXG
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
83QJHY4HS83Q%AXGX981^9#98>T//GJB::83QJHY4HS83Q%AXGX98
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
C::FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%AXG
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
BWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
A::FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%AXG
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
83QJHY4HS83Q%AXGX981^9#98>T//GJB::83QJHY4HS83Q%AXGX98
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
C::FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%AXG
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
BWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
A::F**ENGLISH**K**IMPACT**//17B=6>**AN****E****V****A****L****U****A****T****I****O****N****O****F****U****E****N****G****L****I****S****H**AXG
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
G%2AXGML1981^9#DRM<L%9M**L****A****N****G****U****A****G****E****Q****C****A****P****A****B****I****L****I****T****Y**%L1UM8P2
83QJHY4HS83Q%AXGX981^9#98>T//GJB::83QJHY4HS83Q%AXGX98
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
?//GWB::16B=6PW8FW?9>?//GWB**M****A****D****R****I****D****,****S****P****A****I****N**4>JHY4HS83QJHY4
C::FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%AXG
G%2AXGML1981^9#DRM<L%9M**O****C****T****O****B****E****R****I****2****0****1****7**535%AXG%L1UM8P2
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
BWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
A::FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%AXG
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
G%2AXGML1981^9#DRM<L%9M**E****L****I****Z****A****B****E****T****H****S****H****E****P****H****E****R****D****G****A****N****D**UM8P2
83QJHY4HS83Q%AXGX981^9#98>T//GJB::83QJHY4HS83Q%AXGX98
FWB::17B=6PW?9BEB17OAGWB::**V****I****C****T****O****R****I****A****W****A****I****N****S****W****O****R****T****H**8FW8FW?9B
?7GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
C::FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%AXG
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
BWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
NOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?9W
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
A::FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%AXG
?>GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
9%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2
83QJHY4HS83Q%AXGX981^9#98>T//GJB::83QJHY4HS83Q%AXGX98
FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?9B
?//GWB::16B=6PW8FW?9>?//GWB::17B=6PW8FW?4>JHY4HS83QJHY4
G%2AXGML1981^9#DRM<L%9M5<ZV28TIQU<2535%AXG%L1UM8P2

FOREWORD	02
CONTRIBUTORS	05
1. EXECUTIVE SUMMARY	06
2. INTRODUCTION	08
3. RESEARCH METHODOLOGY	12
The sample design: a stratified two-stage cluster sample design	12
Method of delivery	16
The English language assessment	16
The student context questionnaire	19
4. IMPLEMENTATION IN MADRID	28
5. LANGUAGE LEARNING ENVIRONMENT	34
6. SAMPLING RESULTS	40
7. ASSESSMENT OUTCOMES	44
8. ENGLISH LANGUAGE LEARNING MOTIVATION	56
Construct validation findings	56
Female and male student populations' motivation	62
Bilingual and non-bilingual schools' student populations' motivation	65
Socio-economic status (SES)	68
Links between motivation and proficiency	73
9. IN CONCLUSION	86
REFERENCES	88
APPENDICES	92

FOREWORD

Andy Mackay

Country Director, British Council, Spain

‘We Spaniards spend our life studying English but we never master it’. Were I to be given one Euro for every time I hear that, I would be a wealthy man. And I don’t agree: having returned to live in Spain after a gap of 15 years, my experience has been that language levels have improved very significantly in the interim. But, up until now, we have not had reliable evidence to support either view.

Spain has been a leading country within Europe in its commitment to the development of CLIL and bilingual education. This has been championed from the outset by the Ministry of Education, Culture and Sport and by the regional Education Departments. In partnership with the Ministry of Education, the British Council started a national bilingual programme in schools twenty years ago. In the Community of Madrid, a programme which began in 26 infant and primary schools in 2004 is in place in 515 centres at the beginning of the 2017–18 school year, making it the largest bilingual programme in Spain.

We feel privileged that the Community of Madrid has been our partner in the development of the English Impact project. The project allows us to understand better the achievements of the Madrid bilingual programme over the thirteen years it has been in place, both in terms of the progress made by the students and of their learning orientation and motivation. What stands out from the results is that students from bilingual schools performed better than students from non-bilingual schools across all skills.

This provides clear evidence that access to quality bilingual education is helping young Spaniards to become more proficient and more confident language learners.

What’s more, the students from the bilingual programme show greater levels of international orientation. This is particularly important for us at the British Council. Our mission is to give young people from different countries, cultures and linguistic backgrounds the opportunity to develop the skills to engage in the global environment and to build relationships with people and peoples around the world.

The combination of higher levels of achievement in English and a more international outlook means that the demand in Spain for the English language, the British education system and for qualifications from the UK is huge. And it continues to grow: more and more Spaniards are taking up opportunities for international education: in 2016, there were almost 11,000 studying at universities in the UK, an increase of 11 per cent in only one year.

At the British Council, we hope that the results of this English Impact research will help to build on the success of the bilingual education programme in the Community of Madrid by providing data and evidence to inform its future development.

And the results mean that we can now challenge, with evidence, the belief that English language levels in Spain are not improving – but improving fast.

Barry O’Sullivan

Head of Assessment Research and Development, British Council

English Impact, perhaps the British Council’s most ambitious language-related research undertaking to date, has the capacity to offer ministries of education an accurate and objective diagnostic of the language capability of their country, region or city. This methodology is particularly powerful as it is based on expert statistical design and analysis combined with contextually appropriate interpretation of data that is a precise representation of the general population for the age-group examined.

We have learned so much in the process of designing and delivering this project that it is clearly impossible to highlight all. The sampling work undertaken with Martin Murphy and his team from the Australian Council of Educational Research (ACER) is the single exception. Their professionalism helped to take our vision to a new and elevated level. Their clear and thoughtful consideration of how comparison units should be defined and how to ensure that the final test population was truly representative brought significant challenges, but also significant improvement to our understanding of the processes involved in complex sampling. The results described in this report are testament to the impact that Martin’s thinking has had on our work.

The reality of delivering English Impact was a challenge that would stretch the project team to its limits. In fact, without the key players who undertook the research design and delivery, it is highly unlikely that I would be writing this foreword. We were incredibly lucky to work with exceptional local British Council and Ministry teams in Madrid who did a great deal to ensure the success of the data collection. The determination of these two teams to deliver the project to the highest possible level of quality was critical to its eventual success.

I must confess to feeling great pride in the completion of the English Impact project in Madrid. This report demonstrates the highest level of professionalism and will come to be recognised as a major achievement both within the British Council and in the world of English language education and policy. I expect that it will help the Ministry to continue to conceive and pursue successful policies for many years to come.

Martin Murphy

Senior Research Fellow, Australian Council for Educational Research

At the heart of all good educational policy and practice are teachers, schools, and school systems working to improve the learning outcomes of students. Every day educators and policy makers globally are striving to achieve this goal in very different circumstances.

Learning from experience is an established method of improving performance. I believe all good teachers learn from their colleagues as all good schools share their experiences with other schools within their system. The same can be said for developing policies and practice at the system level. This is where English Impact aims to contribute high quality international comparative outcomes data on English language learning for this purpose.

Education systems are complex. They are shaped by many factors such as geographic location and social and economic background. By mapping the British Council's Aptis assessment outcomes onto a common population framework and by quantifying national and regional variations against that international framework, English Impact aims to identify educational policies and practices associated with the successful teaching and learning of English.

Australian Council for Educational Research (ACER) has thoroughly enjoyed its collaboration with British Council in this endeavour, helping to bring to English Impact the same methodologies underlying major international surveys such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) in the development of this population framework, sampling, weighting and variance estimation.

ACER congratulates the Ministry of Education of the Madrid Region and the British Council team from Spain for their very successful implementation of English Impact. The quality of the survey implementation - evidenced by the very high rates of participation and coverage, and levels of precision that meet or exceed the standards of TIMSS or PISA – should give every confidence to readers of this report, and those keen to learn from Madrid Region's experiences in the increasingly important field of English language teaching and learning.

CONTRIBUTORS

We would like to acknowledge the invaluable contributions made by British Council staff in Madrid, the consultation provided by the Australian Council for Educational Research and the University of Bath. Without the hard work and expert involvement of the following people, English Impact Madrid would not have been possible.

British Council, Spain for its implementation of English Impact in Madrid:

- Diane Pozuelo, Aptis Co-ordinator
- Maureen McAlinden, Programmes Co-ordinator
- Mark Levy, Head of English Programmes

The **Australian Council for Educational Research** for its specialist contribution to sampling methodology and analysis:

- Martin Murphy, Senior Research Fellow
- Clare Ozolins, Research Fellow
- Kate O'Malley, Research Fellow
- Jorge Fallas, Research Fellow
- Tim Friedman, Research Fellow
- Bethany Davies, Research Officer

The **University of Bath** for its contribution to motivational scale questionnaire design and analysis:

- Dr Janina Iwaniec, Lecturer in TESOL, University of Bath

Dr Karen Dunn, Senior Researcher, British Council, for her contribution to statistical analysis.

Many thanks also to the following for their valued support and input:

Members of the British Council English Impact Board

Cherry Gough	Barry O'Sullivan
Rebecca Hughes	Anna Searle
John Knagg	Greg Selby
Maureen McDonagh	Duncan Wilson

Members of the British Council Assessment Research Group

Vivien Berry	Judith Fairbairn
Victoria Clarke	Barry O'Sullivan
Jamie Dunlea	Mina Patel
Karen Dunn	Richard Spiby

Key members of the British Council Tests Team

Dan Elsworth	Jermaine Prince
Catherine Hughes	

Finally, a special thank-you must go to all the participating schools, their dedicated staff and the Community of Madrid for facilitating the English Impact research in the Madrid region.

1. EXECUTIVE SUMMARY

English Impact aims to provide robust policy-relevant data evidencing English language capability in the Community of Madrid

English Impact aims to assess capability by measuring the current ability of a targeted sample of the school population from government-funded schools in Madrid using an English language assessment, and evaluating potential through an in-depth analysis of students' language learning opportunities in and outside the classroom, their language learning motivations and socio-economic backgrounds.

To ensure the English Impact Madrid data can be used to inform and support education system and policy development, the British Council worked in collaboration with the regional Ministry of Education, tailoring the research to meet local needs. This included a detailed investigation into the comparison between bilingual and non-bilingual schools' performances.

The rationale for undertaking this research is supported by the British Council's Royal Charter and charitable objective to develop a wider knowledge of the English language and looks to build upon the organisation's rich heritage of global English language research. A pioneer of the study of English language, the British Council has significant experience contributing analysis and insight, while advancing knowledge across the field. While previous research has explored and expanded existing understanding of how growth in the use of the English language could shape the world economy, English Impact creates new baseline data to measure levels of English language capability.

The research was carried out by the British Council with contributions from the Australian Council for Educational Research and the University of Bath. In 2016–17, the research was piloted within four regional- and national-level education systems: the Community of Madrid in Spain, the Metropolitan District of Bogotá in Colombia, Sri Lanka and Bangladesh.

English Impact in Madrid

- English Impact employs a two-stage cluster sample design used by recognised large-scale international surveys, sampling schools at the first stage and students at the second stage.
- 170 government-funded schools and 2,028 students were sampled to participate in English Impact 2016–17, with 169 schools and 1,774 students participating following exclusions, student withdrawal or absence.
- In order to compare outcomes from bilingual and non-bilingual schools, bilingual schools were oversampled to achieve sufficient student numbers to make precise comparisons. This resulted in a sample of 125 non-bilingual schools and 45 bilingual schools.
- Students were sampled from compulsory secondary education (ESO) 4. This grade represents ten years of schooling from the first year of ISCED Level 1, and a mean age at the time of testing was at least 15 years six months.

- Students sampled were studying English at ESO 4 grade level. A minimum of 90 minutes of formal English study per week as part of the school programme was required for eligibility to the target population.
- Madrid students completed the British Council's Aptis for Teens English language assessment, which tests reading, writing, speaking, listening, and grammar and vocabulary.
- A questionnaire comprising 51 items, delivered in Spanish, captured opinions and data from students on their school and language learning backgrounds, their language learning motivations and socio-economic status.
- Female students performed better than male students across all skills.
- 89.6 per cent of participating students began learning English in preschool or grade 1.
- Motivation is clearly related to proficiency, with confidence in language learning found to be most closely related to achievement.
- Language learning motivation of male students was found to be more dependent on the external environment than female students, with boys reporting overall lower levels of motivation than girls.
- Learners from non-bilingual schools and from lower socio-economic backgrounds reported lower levels of self-assurance.
- Students reporting lower socio-economic status showed significantly lower reported levels of motivation on all scales.

Key findings

- Overall school and student participation in Madrid exceeded the English Impact international participation standard of at least 85 per cent of sampled students in 85 per cent of sampled schools.
- Students from bilingual schools performed better than students from non-bilingual schools across all skills.
- 34 per cent of the participating population achieving at B2 or C CEFR level in the English language assessment.
- 38.5 per cent of the participating population achieved at B1 CEFR level.
- Across the region, listening was the most highly achieving skill, followed by writing, then reading, with speaking the lowest achieving skill.

2. INTRODUCTION

English language learning plays a significant role in many national and regional education systems, increased proficiency having been identified by policymakers as contributing to economic prosperity. Examples of government policies that prioritise the improvement of English proficiency can be seen across the world, the Community of Madrid being one long-standing and well-known example. Considerably harder to find is good data that provide a comparable baseline of evidence showing levels of English language capability at the heart of where government policy makes an impact – in publicly funded school classrooms.

Highly influential sources of data assessing academic achievement across public education systems do exist in the shape of PISA (the Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study). Collectively known as international large-scale surveys and administered by the Organisation for Economic Co-operation and Development (OECD) and the International Association for the Evaluation of Educational Achievement (IEA) respectively, their results are at the same time eagerly awaited and severely criticised for their deeply influential impact on educational practices in many countries. To date, none have included the assessment of language, but much can be learned from the processes designed to sample and implement large-scale research of this kind from decades of experience.

This global best practice in research, and experience of data collection, is emulated within the design of the English Impact methodology that will be detailed in the following chapters. When designing this research we have also tried to learn from the potentially damaging effect international large-scale surveys can have. By identifying the best and, by default, the worst-performing education systems, international large-scale surveys can, at times, have a negative impact. In anticipation of this perhaps inevitable ‘horse race’, an adaptation of the concept of capability underpins our research design.

The theoretical basis used to define English language capability is derived from an adaptation of Amartya Sen’s capabilities approach. An eminent economist, philosopher and driver of social change, Sen’s revolutionary contribution to development economics involved defining the concept of capability. First conceived in the 1980s as an approach to welfare economics, the theory became predominant as a paradigm for human development, and inspired the creation of the UN’s Human Development Index. Sen describes the capabilities approach to human development as ‘a concentration on freedom to achieve in general and the capabilities to function in particular.’ The core concepts within his theory surround functionings that are explained in relation to achievements, and capabilities as people within societies possessing the opportunity to achieve (Saito, 2003).

Central to Sen's concept of capability was combining functionings and capabilities, achievement and opportunity. This adaptation of English language capability can therefore be described in terms of the level of achievement, or proficiency, reached by a defined population; and the opportunities provided to them to achieve greater proficiency via teaching and learning practice derived from a policy or national guideline.

Achievement, proficiency, progress or aptitude of individual English language learners are most commonly measured by a language test. Bachman (1990) suggests that as research instruments, language tests can support investigations into the nature of language proficiency and language teaching practice and perform a role in programme evaluation only when combined with other forms of data. Critical language-testing theorists also believe the knowledge created via a test is 'narrow and simplistic [...] it is mono-logic based on one instrument which is used on one occasion, detached from a meaningful context.' They suggest that using a test can provide 'a quick fix' (Shohamy, 1998), and an instant solution. However, analysis of data captured via this method alone overlooks the complexities of broader subject matter and is meaningless for the reform of education policy.

This evaluation of English language capability, presented here as an adaptation of Sen's capabilities approach, does therefore not only involve the measurement of English language proficiency captured by a test. Other data was captured and combined to provide full context to our analysis: language policy, language learning environment, language proficiency and language learning motivations. The presentation of this supporting data is intended to provide depth and insight into students' assessment outcomes, and go some way to showing the impact of English language policy in Madrid.

The research aims outlined and investigated were:

- evaluate the English language capability of students studying at state schools within the Madrid region of Spain
- compare the outcomes in schools participating in the region's bilingual programme with those not yet participating in the programme
- understand the relationship between English language learning motivation and increased proficiency.

To achieve these research aims, the British Council brought together world-leading research specialists in collaboration with our own expertise in English language assessment to create the English Impact research methodology.

A :: FNOAG//XCKPW?9BE//17B=6>?BEB17B=ZV28TZQU-<2535%ANOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17? ?//GWBSpain,MADRID9>?//GWB::17B=6PW8FW?4>JHY4H113QJHG%2AXGML1981^9#DRM<L%A9M5<ZV28TIQU<2535%AXG%L1UM8FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?



3. RESEARCH METHODOLOGY

The English Impact research methodology was designed in direct response to the research aims outlined in Chapter 2, to provide the most credible evaluation of English language capability within the Madrid region of Spain. The concept of capability is characterised by the unique combination of understanding both current achievement and future opportunity, by its nature involving analysis of multiple data to capture students' current ability and future potential to succeed. Also fundamental to the evaluation of English language capability within a national or regional education system is an appropriate sampling methodology used to accurately reflect the population of interest and supply sufficiently precise estimates. Reflecting the theoretical framework of capability, the research methodology involves three central components:

- the sample design
- the English language assessment
- the student context questionnaire.

THE SAMPLE DESIGN: A STRATIFIED TWO-STAGE CLUSTER SAMPLE DESIGN

The sampling methodology was designed by the Australian Council for Educational Research (ACER) based on its extensive experience in large-scale international educational surveys. The procedures used were drawn extensively from the practices and experiences of major comparative educational surveys that have been operating internationally for well over a decade, in particular surveys of the International Association for the Evaluation of Educational Achievement (IEA), the Trends in International Mathematics and Science Study (TIMSS) and

the Progress in International Reading Literacy Study (PIRLS), as well as surveys undertaken by the OECD, specifically the Programme for International Student Assessment (PISA). These surveys are highly regarded internationally for their quality, and have become major contributors to educational research and policy development around the world.

The British Council team in Madrid participated in a detailed sampling process designed by ACER and modified locally to ensure all procedures were feasible. An overview of the two-stage cluster sampling activities can be found below:

- Preparation
 - define the comparison unit
 - identify exclusions
 - determine stratification variables
 - obtain database of schools and agree access
 - agree the sample design.
- School sampling
 - select the school sample.
- School liaison and student sampling
 - obtain student data from schools
 - select student sample
 - inform schools of selected students
 - arrange dates for English Impact test participation.
- Data tracking
 - track school participation
 - track student participation.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? - B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 1 3 3 Q J H
G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

The British Council team in Madrid worked directly with the Community of Madrid and the Melbourne-based ACER team to gather the school and student data needed to complete the sampling process described above.

Participation standards

The English Impact research was guided by an established set of participation standards drawn from those used within established international surveys described above. Use of these standards enabled precision around the major estimates of the research, namely English language capability, and to maximise the comparability of outcomes across participating countries.

The following participation standards were applied throughout the sampling implementation and analysis procedures.

Standard 1.1 Students in all schools within the comparison unit – including all educational sub-systems – who meet the criteria documented below, are part of the international target population. Students who meet the international target population are referred to as eligible students.

Standard 1.2 The target population should provide the most exhaustive coverage of students. Any deviation from full coverage of the comparison unit needs to be described and quantified in advance.

Standard 1.3 The total of combined school-level exclusions and within-school exclusions within the comparison unit will be no greater than five per cent of the comparison unit target population.

Standard 1.4 Only students within the comparison unit target population participate in the test.

Standard 1.5 The school sample for English Impact Madrid will be drawn using established and professionally recognised principles of scientific sampling.

Standard 1.6 A minimum of 150 schools will be drawn for English Impact Madrid from the comparison unit.

Standard 1.7 The English Impact Madrid school response rate is at least 85 per cent of sampled schools. If a response rate is below 85 per cent then a pre-determined, systematic use of replacement schools will be implemented.

Standard 1.8 The English Impact Madrid student response rate is at least 85 per cent of all sampled students across responding schools. This response rate includes students from replacement schools.

Standard 1.9 Absent sampled students cannot be replaced by non-sampled students.

The international target population is as follows:

Students within the comparison unit enrolled in the grade that represents ten years of schooling counting from the first year of International Standard Classification of Education (ISCED) Level 1, providing the mean age at the time of testing of at least 15 years six months, and who are currently studying English as part of their studies at this grade level, for a minimum of 90 minutes of formal study per week as part of the school programme.

The international target population is defined to ensure comparability across education systems. It is important that students participating in the survey are at equivalent stages of schooling as well as of comparable age. The naming of grades and the age of entry into formal schooling varies between countries. Therefore, the target grade was aligned across countries to allow for accurate reporting.

The International Standard Classification of Education (ISCED)

UNESCO's ISCED is an internationally recognised classification of the levels of schooling across countries, ranging from pre-primary education (ISCED 0) through to tertiary education (ISCED 6). As with IEA studies such as TIMSS, use of this classification will align the levels of education

within individual countries to a common international framework. ISCED 1 is commonly referred to as primary schooling.

Ten years from the start of ISCED Level 1

Drawn directly from TIMSS, this part of the definition is in recognition that the starting age of students into ISCED 1 varies, with students in some countries beginning primary school at a younger age than in other countries.

90 minutes of formal English per week

This definition means that the survey provides an estimate of English language capability for all Madrid students meeting this definition and studying at least 90 minutes of formal English learning per week, rather than for the entire student population of Madrid.

Comparison unit

The term 'comparison unit', used throughout the description of the English Impact research sample design, is an integral part of the research concept and measurement of English language capability to inform more effective policy development, as described in the report introduction.

Many aspects of educational policy development, such as English language learning, often occur at sub-national levels, e.g. provinces and states. Within provinces or states there may be further divisions – for example between public and private sectors. There is increasing recognition that at national level the focus of international large-scale surveys can be limited with respect to exploring aspects of educational provision that can vary within participating countries.

Where educational provision is primarily the responsibility of provinces or states, using a province or state as the comparison unit can

allow for local policies and practices to be clearly related to results, rather than diluted by a national result where variation in conditions between states or provinces can mask these local effects.

Precedents established in PISA's inclusion of 'adjudicated regions' and TIMSS' use of 'benchmarking entities' alongside national level units of comparison have informed the comparison unit policy implemented throughout English Impact. Close adherence to the participation standards and population definition described above, in combination with concisely described and internationally recognisable units of comparison, informed the decision to allow both national and sub-national comparison units to participate within the research project.

Precision of estimates

The primary basis for the determination of sample sizes is the desired precision of major outcomes from the survey. Common practice is the presentation of this measurement in the form of standard errors and/or confidence intervals around survey estimates. This practice will be followed in presentation of English Impact outcomes. The following minimum sample size for each comparison unit was recommended for every participating comparison unit:

- a minimum of 150 schools
- a target of 12 students from each sampled school
- a target of 1,800 students overall.

Drawing further on established standards used in large-scale international surveys such as TIMSS and PIRLS, thresholds for desired standard errors measurements were established. TIMSS and PIRLS report scores on a scale with a mean of 500 and a standard deviation of 100. To achieve this level of precision, these surveys aim to achieve a sample size such that the standard error is no larger than .035 standard deviation units. This equates to a standard error no larger than 3.5 points. This standard error means a 95 per cent confidence interval of ± 7.0 points around the estimated mean.

For percentage estimates, like the percentage of students in each CEFR level for English Impact, the maximum standard error desired was set at 1.75 per cent of the percentage estimate. This means that the confidence interval around population percentage estimates should be less than ± 3.5 per cent.

Coverage and exclusions

All students enrolled in the target grade studying at least 90 minutes of English per week and within the comparison unit belong to the target population. The target population is intended to provide full coverage of all eligible students within the comparison unit. Any deviation from full coverage of the comparison unit was described and quantified in advance of the data collection phase. Every effort was made to ensure complete coverage of the whole population, however in all established sampling exercises of this kind there are often practical reasons invoked for excluding schools and students:

- school exclusions may include schools that are very remote or very small
- student exclusions include students with either functional or intellectual disabilities that prevent them from taking part in the assessment that fits the predefined criteria.

To ensure comparability and maximum coverage of the eligible population, the standards for English Impact require that school and within-school exclusions should not exceed five per cent.

Stratification

A process of implicit stratification was implemented throughout the English Impact sampling methodology. Implicit stratification has the effect of sorting the school sampling frame by a set of implicit stratification variables. It is an effective way of ensuring a proportional allocation of schools across all implicit strata in the sample. Common stratification variables include urban or rural school status, geographic region or school funding type. Stratification can lead to improved reliability of survey estimates provided the stratification variables are related to those survey outcomes.

METHOD OF DELIVERY

To carry out the assessments in every sampled school in Madrid, a pioneering digital method of delivery was developed. Every English language assessment and student questionnaire was completed by students via an offline-enabled tablet. Other large-scale assessments such as PISA and TIMSS have made initial steps towards computer-based assessment; English Impact has pioneering completion of a large-scale assessment using 100 per cent computer-based delivery.

Data was collected via two applications (apps) on each tablet in fully invigilated conditions. A keyboard was used for the writing component to make this process as easy as possible. Individual headphones with a microphone were used for the speaking and listening components. This delivery method aimed to ensure all students were tested as consistently as possible despite location, internet access or available in-school facilities. Fully computer-based delivery allows like-for-like comparison of results and research outcomes that are robust, reliable and consistent.

The two research tools used to capture data via the tablet apps, the English language assessment and the student context questionnaire, are described below.

THE ENGLISH LANGUAGE ASSESSMENT

The English language assessment instrument used to measure the English proficiency of participants in the English Impact research was the British Council's Aptis for Teens test assessing four skills: reading, writing, speaking and listening, as well as grammar and vocabulary.

The Aptis test system

Aptis is a computer-based test of general English proficiency and has four main variants:

- Aptis General
- Aptis Advanced
- Aptis for Teachers
- Aptis for Teens.

No specific cultural or first language background is required, and test content is developed to be appropriate for English language learners in a variety of contexts. Aptis General, Aptis for Teachers and Aptis Advanced are designed for adults and young adults aged 16 years or over. Aptis for Teens is for 13- to 17-year-olds. An important feature of the tests is their alignment with the Common European Framework of Reference (CEFR), a widely used international framework of language proficiency providing detailed descriptions of what language learners are able to do with a language at six levels of proficiency (Council of Europe, 2001). Incorporating the CEFR in the development of the Aptis test system helps to interpret results by linking the test to an internationally recognised set of proficiency benchmarks.

All Aptis test variants are designed to provide information on the ability of test takers to participate in a wide range of general language use situations. The Aptis test system is an approach to test design, development and delivery that was devised by the British Council to provide flexible English language assessment options to test users. There are five components: core (knowledge of grammar and vocabulary), reading, listening, writing and speaking. Although the core component is always administered, organisations are able to select any combination of the other components according to their needs. For English Impact in Madrid, all five components were taken.

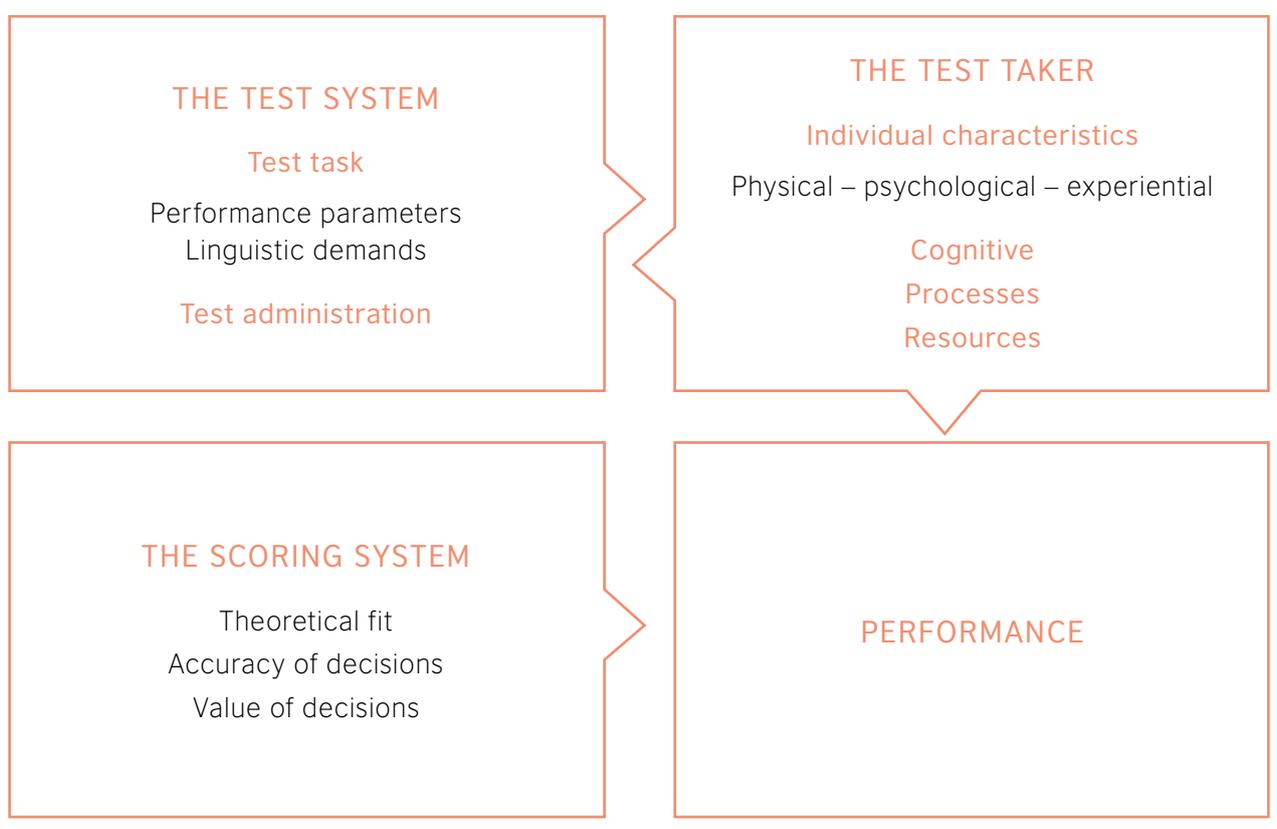
Theoretical model underpinning the test system

The theoretical model of test development and validation that underpins the Aptis test system is based on the socio-cognitive model proposed by O’Sullivan (2011, 2015), O’Sullivan and Weir (2011), and Weir (2005). As O’Sullivan (2015) notes: ‘the real strength of this model of validation is that it comprehensively defines each of its elements with sufficient detail as to make the model operational.’ The socio-cognitive model is based around three elements:

- the test taker
- the test system
- the scoring system.

The model drives design decisions by specifying how these three elements combine to result in a measure of candidate performance that is meaningful in terms of the English language ability being assessed. This in turn allows the test developers to collect evidence in a systematic way in the creation of a validation argument to support claims about the test. Figure 1, taken from O’Sullivan (2015), demonstrates how the three elements feed into the test takers’ performance.

Figure 1: The socio-cognitive model for test design and validation



Research evidence supporting the validity of the test system

An important part of the Aptis test system has been the commitment of the British Council to support an active and robust validation research and dissemination agenda. A dedicated team carries out research and statistical analyses at the design and development stage. Operational test delivery data is regularly analysed to ensure the tests function to demanding technical performance criteria. The Assessment Research Awards and Grants scheme funds research into the tests from leading international researchers. An impressive body of published documentation covering an extensive and diverse range of validation projects contribute important evidence to the validity argument supporting the uses of the Aptis test system.

Localisation: adapting tests for particular uses

The term 'localisation' is used within the Aptis test system to refer to the ways in which the Aptis test is adapted for use in particular contexts with particular populations to allow for particular decisions to be made. The model identifies levels of localisation depending on the degree of change from the original underlying framework used in the development of Aptis, and the amount of resources required to realise that change. Aptis for Teens is considered to be a level-four localisation based on the five-level model described in O'Sullivan and Dunlea (2015). The description for level four is reproduced in Table 1.

Table 1: Level of localisation for Aptis for Teens (from O'Sullivan and Dunlea, 2015)

Level 4	Partial re-definition of target construct from existing variants. Will involve developing different task types to elicit different aspects of performance.	Developing new task types that are more relevant for a specific population of test-takers, while remaining within the overall framework of the Aptis test system (e.g. Aptis for Teens).
----------------	--	--

The Aptis for Teens test

The Aptis for Teens test variant used in the English Impact project has been designed specifically to meet the needs of younger language learners by testing their English language skills through familiar scenarios. Task parameters such as topic, genre and the intended audience are relevant to the target use domain of a teenager. Questions reflect activities that occur in everyday life such as social media, homework, school events and sport. For example, instead of writing a complaint letter to a company – a task used in Aptis General for adults but something a teenager may not yet have experienced – they might be asked to write about the benefits and drawbacks of a social issue relevant to teenagers and likely to be discussed in classrooms. The cognitive competences of the age group are also taken into consideration.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? - B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 1 9 3 Q J H
G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Tasks are tailored to provide support needed for this age group to give them the chance to perform to the best of their ability. For an overview of the structure of each component of the Aptis for Teens test, see Appendix A.

Aptis for Teens was designed for a specific age group – young learners aged from 13 to 17. As with the other main variants in the Aptis test system, the test is designed to be used with test takers irrespective of culture, country of origin or residence, gender or first language. This means that background knowledge is not tested, bias is reduced and language skills are isolated for testing.

An important part of the features impacting on the test system also relates to the test delivery environment. The English Impact project tests were invigilated by a British Council employee, who visited each school to conduct the testing. This additional level of quality assurance ensured the security and uniformity of the test delivery.

Scoring and reporting

The scoring system is the final area of validation. The core, reading and listening components are scored automatically within the computer delivery system. Trained raters mark the speaking and writing components using an online rating system.

Aptis for Teens test results are reported on a numerical scale (0–50) and as a CEFR level for each component. An overall CEFR level is also given if all components are completed by the test taker. The CEFR level describes English language proficiency across six levels (A1–C2). In Aptis for Teens, results are reported for levels A1 to B2, and if a test-taker demonstrates an ability beyond B2, this is reported as C (C1 and C2 are not differentiated in Aptis for Teens).

The core, reading and listening components use selected-response formats such as multiple choice, gap fill and matching tasks. Speaking and writing components require test-takers to provide samples of spoken and written performance. The speaking test is a semi-direct test in which test-takers record responses to pre-recorded prompts. The writing test approximates online written communication. The focus of the speaking and writing marking scales is on test-taker communicative competence and is marked by trained raters. See Appendix A for a detailed overview of the task types contained in each component.

THE STUDENT CONTEXT QUESTIONNAIRE

Context questionnaires form an integral part of most large-scale surveys. While accurate information on student performance is central to such assessments, the factors that are linked to performance are of particular interest not only for researchers but also for practitioners and policymakers in education.

Alongside the four-skill English language assessment, students complete a background questionnaire to gather contextual information to support the English language capability data.

The student questionnaire comprises 51 items in three sections.

Demographic background, including socio-economic status variables

Items within this section of the questionnaire include grade, gender, age, prior schooling, and language spoken within the home and country of birth. A number of questionnaire items were used to measure the latent variables of socio-economic status (SES). This will be explained and explored fully in Chapter 8.

English language learning inside and outside school

Items within this section of the questionnaire include the grade at which students began learning English, time spent learning English and whether students study English outside of school.

English language learning motivation

This section of the student-context questionnaire was designed by Dr Janina Iwaniec from the University of Bath, a second-language learning motivational specialist who conducted a review of the most relevant and influential theories and constructs used to measure language learning motivation. Motivation is one of the most influential of all individual differences, trumping factors such as language learning aptitude (Gardner and Lambert, 1972) in explaining gains in proficiency in certain contexts. Recently, it has been shown that motivation is more important than the age of learning onset, with students who start later developing higher levels of motivation and quickly catching up with the proficiency of learners who started English instruction early (Pfenninger and Singleton, 2016). Motivation is also considered to be responsive to appropriate interventions (Taylor and Marsden, 2014) and can be enhanced or decreased as a result of language learning environment (Ushioda, 2009). This relatively strong influence on language learning and its malleability make motivation a factor that is crucial for language learning policies.

In the years since research into language learning motivation started in the 1950s (Gardner and Lambert, 1959), there have been a large number of theories of language learning motivation proposed. The choice of constructs for English impact was guided by the most up-to-date theories and research on language learning motivation. This included the L2 Motivational Self System (Dörnyei and Ushioda, 2009) which consists of three constructs: ideal L2 self, ought-to L2 self and language learning experience; international orientation (Yashima, 2009); and self-concept (Bong and Skaalvik, 2003).

The scales were measured by 32 questions addressing eight areas of language learning motivation. These hypothesised motivational scales have strong theoretical and empirical grounding described in Table 2. Including four questions for each scale increases the value of the information gathered for each area – something that is exploited in the analytic approach. Grouped in one section of the questionnaire and in a random order, students were asked to give a response to each statement using a six-point Likert scale.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 2 1 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Table 2: Reference table of motivational scales and associated descriptions

Motivational scale (latent variable)	Descriptive name	Motivational scale descriptions	Question reference	Question details
Ideal L2 self (IDEAL)	<i>Personal language goals</i>	<p>Ideal L2 self is an image of oneself as a proficient speaker of a second language (Dörnyei, 2005). Though it relates to the future-self, <i>ideal L2 self</i> needs to be considered attainable to retain its motivational properties. English Impact employed the Iwaniec (2014) scale as it encompasses the four skills of language learning.</p>	11	I imagine myself speaking English fluently.
			12	I imagine myself comfortably reading in English on the internet.
			13	I imagine myself easily being able to follow what others say to me in English.
			14	I imagine myself writing emails in English with ease.
Ought-to L2 self (OUGHT)	<i>Social expectations</i>	<p>The ought-to L2 self is based on the expectations placed on students and relates to the ‘attributes that one believes one ought to possess in order to avoid possible negative outcomes’ (Dörnyei, 2005, pages 105–106).</p>	01	I consider learning English important because the people I respect think I should do it.
			02	Studying English is important to me because other people will respect me more if I have knowledge of English.
			03	Studying English is important to me because an educated person is supposed to be able to speak English.
			04	Learning English is necessary because people surrounding me expect me to do so.

Motivational scale (latent variable)	Descriptive name	Motivational scale descriptions	Question reference	Question details
Language learning experience (EXPER)	<i>Interest in learning English</i>	Language learning experience is concerned with the influence of the immediate environment on language learning (Dörnyei, 2005) and implies a strong focus on language learning attitudes.	EX1	Learning English is really great.
			EX2	I look forward to my English classes.
			EX3	I find learning English really interesting.
			EX4	I really enjoy learning English.
Instrumentality (INSTR)	<i>Future opportunities</i>	Instrumentality represents motivation stemming from the practical benefits of language (Gardner and Lambert, 1972). This scale measures the perceptions of usefulness of English on job markets and future prospects.	INSTR1	I need English for my future career.
			INSTR2	The things I want to do in the future require me to use English.
			INSTR3	I study English because it will facilitate my job hunt in the future.
			INSTR4	I study English as it will help me to earn good money.
International orientation (INTER)	<i>Global communication</i>	International orientation is a construct recently developed in response to the changing role of English. It denotes an 'interest in foreign or international affairs [...] readiness to interact with intercultural partners' (Yashima, 2000, page 57). The scale used, found in Iwaniec (2014), lends itself to adaptations that take into account the growth of online interaction rather than travelling abroad.	INTOR1	Studying English will help me understand people from other countries.
			INTOR2	In the future, I would really like to communicate with people from other countries.
			INTOR3	In the future, I would really like to communicate with people from other countries online.
			INTOR4	If I could speak English well, I could get to know more people from other countries via the internet.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 2 3 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Motivational scale (latent variable)	Descriptive name	Motivational scale descriptions	Question reference	Question details
English self-concept (SELF)	<i>Self-confidence in English</i>	Self-concept is 'a person's perception of himself' (Shavelson, Hubner and Stanton, 1976) and this scale relates to self-evaluation in the students' ability to study English. The most common measurement of self-concept is the Academic Self-Description Questionnaire (Marsh, 1990), adapted to language learning by Iwaniec (2014).	SELF1	I usually get good marks in English.
			SELF2	Compared to other students, I'm good at English.
			SELF3	I have always done well in English.
			SELF4	Studying English comes easy to me.
Parental encouragement (PAREN)	<i>Family expectations</i>	Like the ought-to L2 self, parental encouragement focuses on external expectation. As the participants in English Impact are as young as 15 years six months, there is a potential for their motivation to be influenced by their parents or guardians. Parents are considered to be one of the three groups of important others, together with teachers and peers (Williams and Burden, 1997).	PAR1	My parents think I need to know English to be well educated.
			PAR2	My parents have stressed the importance English will have for me in the future.
			PAR3	My parents feel that it is very important for me to learn English.
			PAR4	My parents encourage me to practise my English as much as possible.
Motivated learning behaviour (MOTIV)	<i>Level of effort</i>	Motivated learning behaviour attempts to measure the behavioural component of motivation, i.e. the reported amount of effort a student invests in English language learning.	MB1	I work hard at learning English.
			MB2	I think I'm doing my best to learn English.
			MB3	I put a lot of effort into learning English.
			MB4	I spend lots of time studying English.

Motivational scale analysis

The questionnaire responses were analysed using confirmatory factor analysis (CFA) in Mplus 7 (Muthén and Muthén, 1998–2017). Factor analytic techniques are of great value in motivational research since they explicitly address the requirement to gain insights into constructs that are not directly observable. These abstract concepts are termed latent variables. Examples include depression in psychology, consumer expectation in economics and anomie in sociology. Socio-economic status is also often treated as a latent variable.

In factor analysis, a number of observed variables, or measures, are hypothesised to be indicators of the existence of an underlying latent variable. No individual observed as variable is considered to give a precise measure of a hypothesised latent variable, as there will always be a certain amount of error in the measurement of any observed variable. Much as, for example, a psychiatrist would expect to see high levels of a range of indicators before diagnosing a patient, factor analysis will combine information from a number of observed variables in order to give information about the levels of a hypothesised latent variable.

In the motivational questionnaire employed in the current study, the aim was to capture information about students' motivational levels in eight areas using 32 questions (see Table 2 above for details). Each block of four questions was posed to gain insight into a distinct underlying motivational construct.

The questions were presented to the students in a random order so that as respondents they would not immediately perceive the shared focus of a set of questions. Employing CFA techniques means that the information from all four observed variables can be combined to give a more robust insight into the motivational traits of interest than would be achieved from modelling individual observed responses.

There are two distinct stages of analysis reported here, both of which employ factor analytic techniques:

- *construct validation* – this involves checking whether the questionnaire functioned as expected with regard to gaining insights into different areas of motivation
- *multi-group analysis* – for the purposes of the current report, three sets of comparative student groupings were examined: male/female, bilingual/non-bilingual school attendance, and socio-economic status (based on parental employment status and household possessions as reported by the students).

The scope of the questionnaire analysis reported under these headings is limited to investigating the motivational scales in their own terms. The findings from this initial analysis are then taken forward to investigate the relationships between motivational scales and proficiency as measured by the Aptis test, as described in the next section.

Construct validation

The primary aim in construct validation was to establish whether it makes sense to understand observed student responses with reference to the eight hypothesised motivational scales. In CFA, the measurement model refers to this relationship between the responses given by the participants to the questions (the observed data) and the motivational scales (latent variables) that reflect each of the motivational constructs. The CFA technique can be employed to assess how much of the variation in the original observed dataset can be explained with reference to this pre-defined latent structure. This involves accounting for the shared variation, or correlations, between the observed measures.

Estimates from the measurement model (known as factor loadings) give an indication of how much variation in the observed variable is accounted for by the latent construct. Some variables will have a stronger relationship than others. If there is a close relationship between all observed variables and the associated latent variable, there is a strong internal consistency in the scale. However, it is worth noting that a latent variable will never account for all the variations in a given observed variable – there will always be some measurement error. This reflects the principle inherent to factor analysis in that any given observed variable is driven by an underlying trait (in this case of motivation) and will not provide a precise measure of it. Measurement error takes into account, for example, idiosyncratic responses to questions worded in a certain way.

CFA is a data reduction technique that draws on a reduced number of variables to replicate patterns in the observed data. In order to assess whether the hypothesised measurement model achieves this successfully, a number of fit statistics are employed. These indices represent several different means of indicating how well the latent structure can be used to replicate the variation in the observed dataset. Essentially, if the structure hypothesised by the model is able to capture the patterns of question responses well, the model is considered to be a good fit. 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 (ideally 0.95), and the root mean square error of approximation (RMSEA) below 0.05 for good model fit.¹ Where alterations are made to the initially hypothesised model, comparisons are made using the adjusted chi-square difference test (Satorra and Bentler, 2010). This is the recommended means of comparing CFA models estimated using the maximum likelihood with robust standard errors (MLR) approach² employed in the current analysis (Muthén and Muthén, 2011). Model comparisons are undertaken in a systematic manner to ensure that any additional parameters included in the model bring about an overall improvement to the model fit.

The first step in the analysis reported here is thus to establish a measurement model that reflects the data well. This is carried out for all cases in the dataset together, before moving on to the group comparisons.

1. See Byrne (2012, pages 69–77) for a description of what these indices represent and the values accepted to show good fit.
2. MLR refers to maximum likelihood parameter estimates with standard errors and a chi-square test statistic (when applicable) that are robust to non-normality and non-independence of observation (Muthén and Muthén, 2011, page 533).

Multi-group analysis

There are a number of approaches that can be taken to making multi-group comparisons within a CFA modelling approach (see, e.g., Byrne, 2012, pages 193–281). For current purposes, the focus is on two areas:

- A. comparing relative levels of motivation expressed for each scale
- B. comparing the relationships between the different areas of motivation.

Under (A), the model is used to derive what are known as factor scores for each of the students on each of the motivational scales. In other words, for each of the motivational areas listed in Table 2, each participating student will be assigned a value (factor score) depending on their responses to the relevant question. This is more complex than simply averaging the responses, as it takes into account the weighted relationships estimated within the model. Factor scores are expressed on a standardised scale, which does not bear any easily perceptible relationship to the original measurement scale. However, it is the comparison between levels that are of interest here, and therefore the scale is not essential.

For (B), the relationships examined are those between the latent variables. This relationship is referred as the structural model. In CFA terminology, this is restricted to covariances, which do not presume any directionality in the relationship between variables. The value of investigating these relationships as part of this multi-group analysis is that they tell us whether

the balance between the motivational scales is consistent between groups. The model is set up so that the measurement model is kept consistent across groups, but relationships between latent variables are allowed to vary where significant differences are found.

This is achieved by creating two models: one in which all estimated parameters are assumed to be the same between groups (the constrained model) and one in which all parameters are freely estimated between groups (the configural model). If there is a significant difference between these two models, as assessed by comparing model chi-squares, this indicates some group differences. The aim is then to bring the constrained model closer to the configural model by freeing up parameters of interest, in this case covariances between latent variables. Modification indices (MIs) given by the software show where the greatest improvements in model fit can be achieved. Where there is a significant difference between groups, it will improve the model to estimate the covariances indicated by the MIs separately across groups, rather than constraining them to be the same. Parameters are freed up in a step-by-step basis, and only incorporated in the model if they lead to a significant improvement in the adjusted model chi-square.

This process will lead to a model that has the same latent structure across groups, but for which some parameters (in this case covariances between latent variables) are estimated separately. This enables key differences between groups to be explored.

Links between motivation and proficiency

To explore the relationship between motivational variables and proficiency, two key pieces of information were used to run profile and correlation analysis in SPSS 22 (IBM Corp., 2013):³

- scores from the Aptis test, both overall and for each test component
- factor scores (F-scores) for each participant for each motivational scale.

Aptis test outcomes are available as CEFR levels, as well as scale scores for both individual components and the overall test. The CEFR levels assign participants into a broad proficiency banding, while the scale scores provide a more detailed insight into test performance. In the analysis carried out, the CEFR bands are used to set up comparative groups of students, while the scale scores are used for the more detailed correlational analysis.

The F-scores, meanwhile, were derived from the multi-group CFA described above and ascribe each questionnaire respondent a level for each motivational scale. Essentially, once the CFA model is set up satisfactorily, a value to reflect participants' levels for each latent variable is calculated. So, for example, if a given participant gave strongly positive responses to the questions on English self-concept, they would have a higher fscore for this motivational scale than a respondent who provided low or mixed responses to the same questions. The exact balance of the relationship between observed responses and F-scores is determined by the factor loadings estimated in the model. The F-scores themselves run on a standardised zero-centred continuous scale (between -1 and 1).

3. P-values are derived from a comparison of the correlation co-efficients following Fisher's z-transformation. Calculations performed using <http://vassarstats.net/rdiff.html>

4. IMPLEMENTATION IN MADRID

The main objective of the sample design for English Impact is to present the most accurate possible results based on the comparison unit chosen: state schools within the Madrid region of Spain. In order to meet the established participation standards described in Chapter 3, a precisely defined comparison unit had to be outlined and agreed.

The education system in Madrid

The Madrid region of Spain has an approximate population of 6.5 million. Education is compulsory for ten years and is free in state-funded schools. It is divided into three stages:

- preschool (educación infantil) education from the ages of three to six is non-compulsory in the Madrid region. However, almost 100 per cent of pupils attend preschool. Preschool education in Madrid is most often provided within Primary schools
- primary education (educación primaria), provided in primary schools and covering six years, between the ages of six and 12
- lower secondary education (educación secundaria obligatoria), provided in secondary schools, and covering four years between the ages of 12 and 16.

Types of school

There are 702 high schools in the Madrid area, of which 392 are charter schools and 310 are state schools.

State schools are described as non-denominational schools fully funded by the regional government.

Charter schools are described as private schools receiving a regional government subsidy to provide education services.

This academic year, the number of students registered in ESO (compulsory secondary education) was 258,891. Some 42 per cent of Madrid students attend schools in the city and 58 per cent in its environs. All the students in ESO 4 study English as a second language. Bilingual schools' students study one-third of their subjects in English.

Bilingual schools are government-funded schools that offer bilingual education. There are two streams: sección, where students study three or more subjects in English, and programa, in which pupils study one or two subjects in English.

The organisation and governance of education is the joint responsibility of the Ministry of Education, Culture and Sport (MECD) at national level and the education authorities of Spain's 17 autonomous communities (ministries or departments for education). The MECD executes the government's general guidelines on education policy and regulates the basic elements or aspects of the system. Regional education authorities in the autonomous communities, in this case Madrid, develop the national regulations and have executive and administrative responsibilities for managing the education systems in their own territories, along with local authorities and the educational institutions themselves.

The bilingual programme in Madrid

Although central government has overall responsibility for the education system and regulates the content of curricula and minimum teaching time for core subjects, decentralisation and devolution of responsibility for education to the autonomous communities means they have freedom in educational decision making in general and more specifically in terms of language learning.

In 2004, the Community of Madrid introduced the Bilingual Spanish–English Programme into 26 state primary schools in the region. The most recent statistics show that as of the academic year 2015–16, the bilingual model was implemented across 353 state primary schools and 110 state secondary schools, with further growth expected in 2016–17. There are also an additional 181 state-funded primary schools that offer bilingual education at secondary level in 33 schools.

The bilingual programme is made up of key components that include:

- **staff support** – teachers and programme co-ordinators, principals and language assistants must collaborate to ensure the bilingual programme is implemented to defined standards across state schools in Madrid
- **teacher training** – investment in teacher training is a key commitment by the Regional Ministry of Education. The immersive English Language Teacher Training Programme and the Professional Development of Teaching Competences are offered alongside other courses to maintain the expected teaching calibre
- **certification** – teachers in both state primary and secondary schools must achieve the Linguistic Certificate to fill bilingual teaching roles; they must also teach a required number of hours to maintain certification
- **evaluation** – to ensure continued quality, bilingual programme students take part in external evaluations in year four and six at state primary and year four of compulsory state secondary education.

The initiation of the bilingual programme was the Madrid region's direct response to the European Union's call for member states to prioritise foreign languages, with a specific urge for students to learn two foreign languages by the time they complete their compulsory state education. The aim of using English as a medium of instruction across the curriculum is to increase students' chances of acquiring the language via a more communicative and natural process.

For over a decade, the importance and recognition of the bilingual programme has grown, and through English Impact the opportunity to establish a baseline of performance and compare language capability in bilingual and non-bilingual schools has been utilised.

Sampling bilingual schools to allow comparison

In order to accurately compare the English language capabilities of students across bilingual and non-bilingual schools in Madrid, the sample size for schools was increased from 150 to 170 in order to include an oversample of bilingual schools. The number of schools in the bilingual school stratum was calculated based

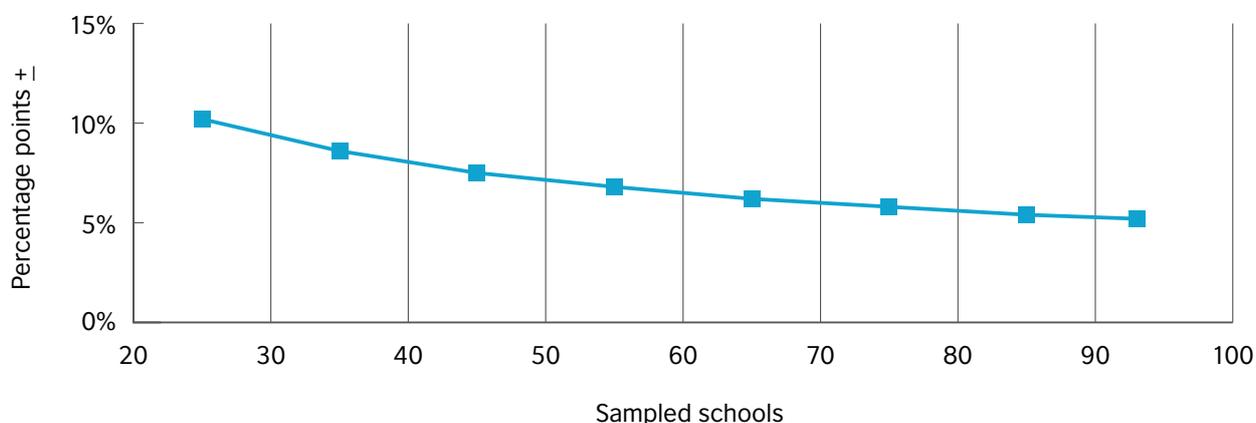
on expected achieved sample size and outcomes from Spain's PISA 2009 results.

Without oversampling, a sub-population such as bilingual students would usually be represented in proportion to its presence in the total Madrid population. However, this would result in a smaller number of bilingual schools than would be required to meet the desired level of precision for estimates for this sub-population. Prior to oversampling, for the Madrid region as a whole confidence interval widths were estimated to be within ± 4.1 per cent; this resulted in confidence interval widths for the bilingual group of ± 10.1 per cent and for the non-bilingual group of ± 4.5 per cent. Reporting confidence intervals wider than ± 10 per cent is generally avoided because the sample estimates become markedly less meaningful.

Increasing the sample size of bilingual programme students

Figure 2 shows the estimated effect of increasing sample size on the size of the confidence interval to be achieved around estimates for bilingual students using the minimum expected sample size of 25 schools, up to a full census (93 schools) of bilingual schools.

Figure 2: Effect of sample size on confidence interval estimates



Estimates of the 95 per cent confidence interval according to sample size for bilingual schools in the region of Madrid – publicly funded schools.

Table 3: The effects of school and student sample size on confidence interval estimates

Sampled schools	25	35	45	55	65	75	85	93
Sampled students	250	350	450	550	650	750	850	930
95% CI	10.2	8.6	7.5	6.8	6.2	5.8	5.4	5.2

As shown in Table 3, accuracy of confidence interval estimates increases with the increase in sample size. Sampling 45 bilingual schools and increasing the overall school sample size to 170 reduces the confidence interval widths for this subpopulation to ± 7.5 per cent, similar to those expected, for example, for results for larger states within Australia in large-scale surveys.

The comparison unit

- The Madrid region of Spain’s government-funded schools.
- Bilingual schools were over-sampled to allow the comparison of English language capability between schools participating in the bilingual programme and non-bilingual schools.

- 170 schools and 2,040 students were sampled, broken down into 125 non-bilingual schools and 45 bilingual schools.

All bilingual schools within the evaluation are public schools as charter schools during the 2016–17 academic year did not have the Comunidad de Madrid’s Bilingual Programme in the 4° ESO age group.

Bilingual schools were identified as an explicit stratification variable and were used to select the school sample, as described above, alongside the implicit stratification variables chosen. All were used to ensure a proportional sample allocation across the implicit strata. The stratification variables applied to the Madrid sample frame are outlined in Table 4.

Table 4: Stratification variables chosen

Stratification variable name	Variable labels
Bilingual education	Bilingual schools to non-bilingual schools*
School type	Charter schools/state schools
Geographic region	North/South/East/West/capital

* Non-bilingual: public and charter schools



ENGLISH

5. LANGUAGE LEARNING ENVIRONMENT

All 1,774 students participating in English Impact Madrid completed a 51-item questionnaire delivered in Castilian Spanish, as described in Chapter 3. Answers reported to 12 questionnaire items are presented Table 5 as a demographic profile.

The gender ratio of female and male students was reported as an almost equal split: 50.9 per cent and 49.1 per cent respectively. As described in Chapter 3, the average age of the target population was 15 years six months. The mean age reported by participating students was 15 years six months, showing the accuracy of students sampled from the target population. The largest proportion, 93.6 per cent, indicated they spoke Castilian Spanish most often at home. When asked to report their country of birth, 84.4 per cent selected Spain;

almost 15 per cent reported a number of other locations as their country of birth, a significant proportion of which were Spanish-speaking: Ecuador, Peru, Colombia, etc. A small number of students (0.7 per cent) reported they most often spoke English at home; these students reported their country of birth as Spain, Canada and the UK.

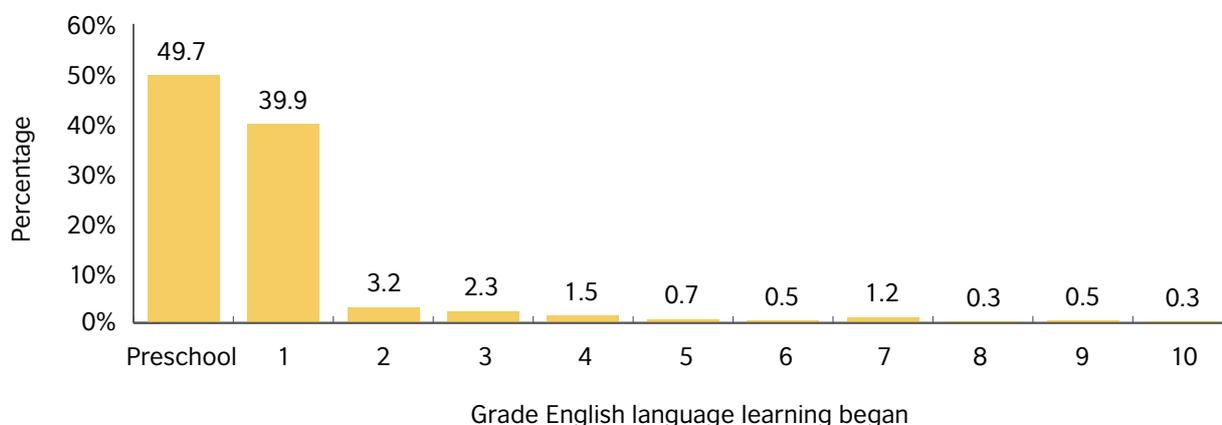
All participating students were asked to report when they started learning English. Some 49.7 per cent indicated they started learning English in preschool with 39.9 per cent starting to learn English in the first grade. Combined, this shows that 89.6 per cent of participating students began learning English before the second grade of their schooling, approximately under the age of nine.

Table 5: Demographic variables of participating students from the Madrid region

Demographic variables	Reported by participating students in Madrid
Gender	50.9% female; 49.1% male
Age	Mean age of 15.6 years
Language most often spoken at home	93.6% Castilian Spanish; 5.6% other; 0.7% English
Country of birth	84.4% Spain; 2.3% Ecuador; 1.9% Romania; 1.2% Peru; 1% Colombia; 1% Morocco; 0.8% China; 0.7% Venezuela; 0.6% Argentina; 6.1% other (less than 0.5% individually)
Attendance at preschool	90.6% attended preschool; 9.4% didn't attend preschool

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 3 5 3 Q J H
G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Figure 3: The grade participating students from Madrid started learning English

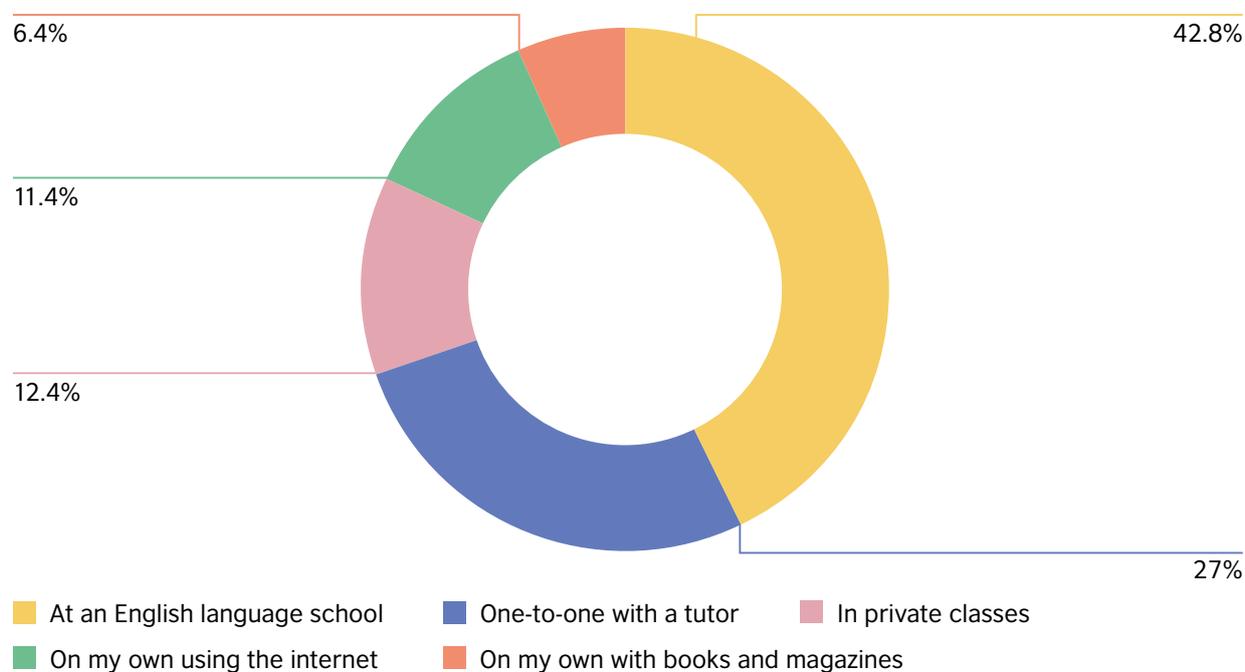


Participating students were asked to provide information on studying English outside of school; 70 per cent said they did not study English outside of school; 30 per cent said they did choose to study English outside of school. Of the 30 per cent of students from the total population that said they chose to study English outside of school, 76.5 per cent were attending non-bilingual schools.

Those students who indicated they did choose to study English outside of school were asked to indicate what activity best described their extracurricular language learning. The largest proportion, 42.8 per cent, suggested they learned at an English language school; 27 per cent with a one-to-one tutor; 12.4 per cent in private classes; 11.4 per cent on their own using the internet; and 6.4 per cent on their own using books and magazines.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W 3 6 1 6 B E N G L I S H ? I M P A C T B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H S 8 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

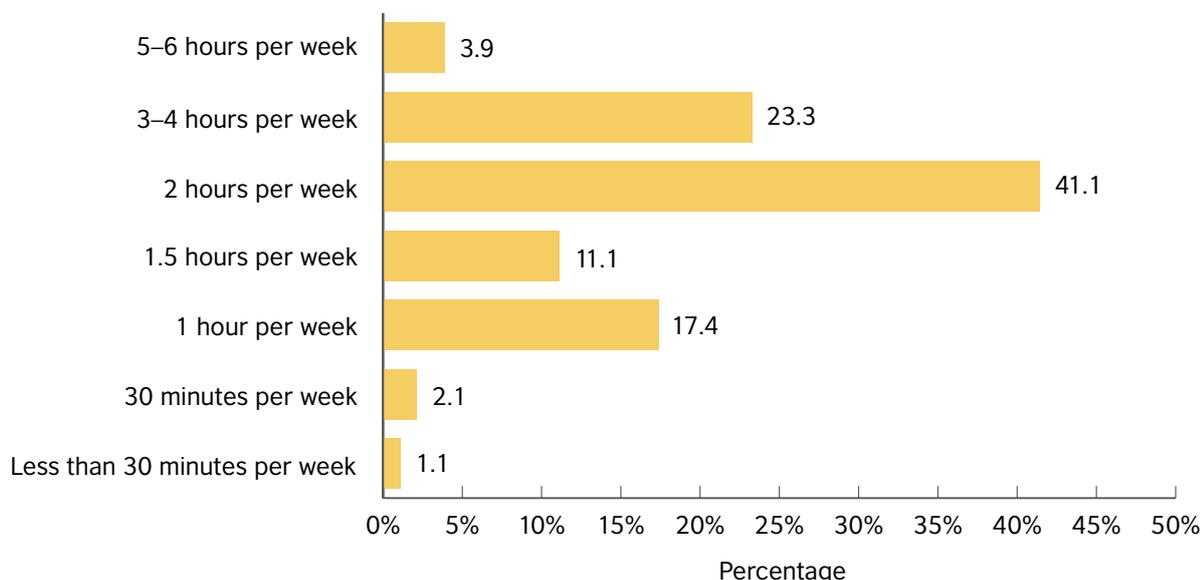
Figure 4: Activity undertaken when studying English outside of school



The same segment of the participating student population were asked to indicate how much time they spent learning English outside of school. On average, students suggested they spent between two to four hours a week on

English language learning additional to regular school hours. The largest proportion, 41.1 per cent, indicated they studied for two additional hours each week.

Figure 5: Time spent learning English outside of school

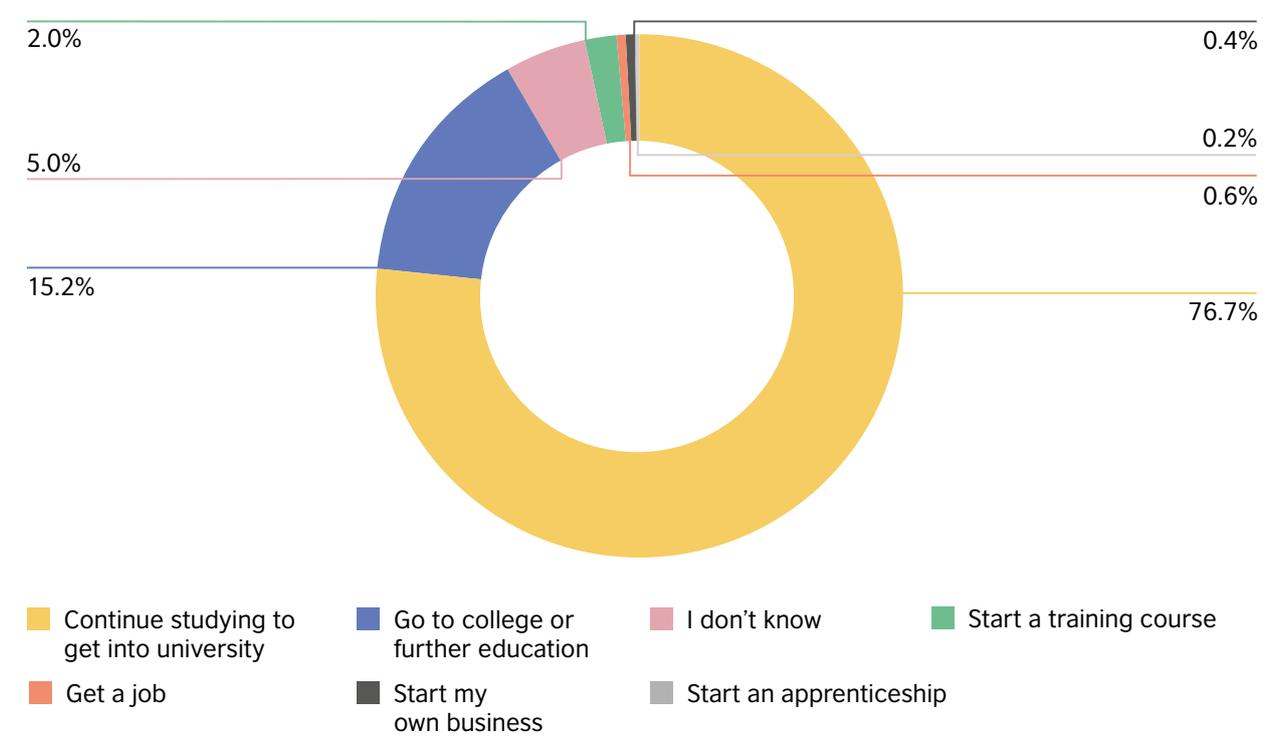


A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 3 7 3 Q J H
G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

To understand more fully the possible further pathways of school students from Madrid, all participating students were asked to report what they would like to do when they finished school. Over three-quarters, 76.7 per cent, suggested they wanted to continue studying to get into university; 15.2 per cent to go to college or into

further education; five per cent said they didn't know; two per cent said they would like to start a training course; 0.6 per cent said they'd like to get a job when they finished school; 0.4 per cent to start their own business and 0.2 per cent to start an apprenticeship.

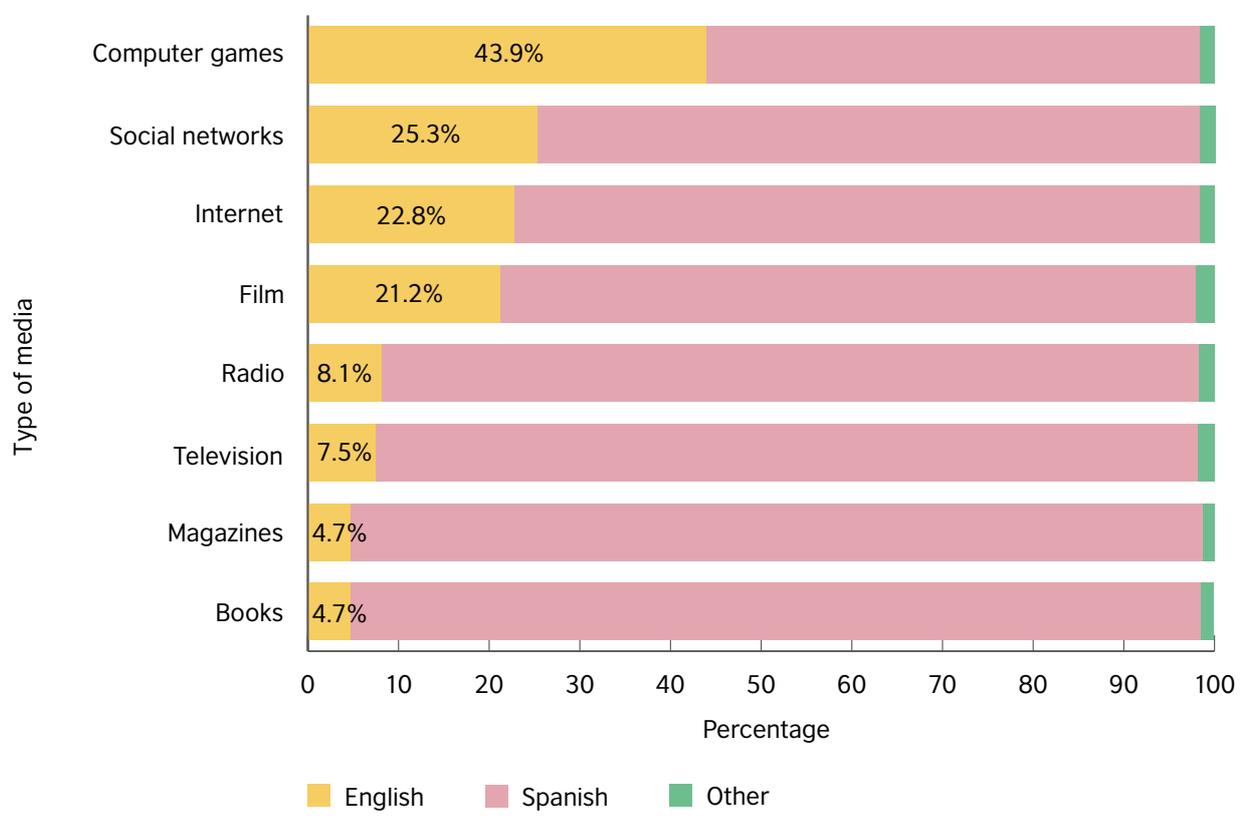
Figure 6: Future pathway when finishing compulsory schooling



The impact of media usage on language learning is often identified anecdotally by teachers and policymakers as a catalyst for accelerated proficiency, especially among the digital generation of millennial learners that comprise the target population of English Impact Madrid. To gauge and understand their language use when interacting with various types of media, they were asked to indicate how they watch, read or listen to the following things most regularly, in English, Castilian Spanish or another language.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W 3 8 1 6 B E N G L I S H ? I M P A C T B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H S 8 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Figure 7: Language most regularly used to watch, read or listen to different types of media



Most notably, 43.9 per cent of participating students suggested they interacted with computer games in English. Of that proportion of the total population that indicated they play computer games in English, 56.4 per cent were male and 43.6 per cent were female. Of the 54.5 per cent of the total population that indicated they played computer games in Castilian Spanish, 56.9 per cent were female and 43.1 per cent were male.

Of students that suggested they use social networks in English, 56.9 per cent were female. A slightly larger proportion of females, 53 per cent, suggested they use the internet in English, and 66.2 per cent of students that indicated they watch films in English were female.

A :: FNOAG // XCKPW ? 9 BE // 17 B = 6 > ? BEB 17 B = ZV 28 TZQU - < 2535 % A
NOAGWBN7 = ? 9WNOAGWB :: E = ? 9WB8FW ? 9BEB 17B98 1% > ? 9BEB 17 ?
? // GWB SPAIN , MADRID 9 > ? // GWB :: 17 B = 6 PW8FW ? 4 > JHY4H39 3Q JH
G % 2 AXGML 198 1 ^ 9 # DRM < L % A9M5 < ZV 28 TIQU < 2535 % AXG % L 1UM8
FWB :: 17 B = 6 PW ? 9 BEB 17 OAGWB :: 7 = ? 9 = W8FW8FW ? 9 BEBW8FW8FW ?

WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
81X9 # 98 > T // GWB :: NOAGWB :: 7 = ? 9WNOAGWB :: E = ? 9WB8FW ? 9B7B1
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
GML 198 1 ^ 9 # DHM < L % A9 :: 5 < ZV 28 TIQU < 2535 % AXG % L 1UM8 :: 2 ^ Q
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
81X9 # 98 > T // GWB :: NOAGWB :: 7 = ? 9WNOAGWB :: E = ? 9WB8FW ? 9B7B1
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
GML 198 1 ^ 9 # DHM < L % A9 :: 5 < ZV 28 TIQU < 2535 % AXG % L 1UM8 :: 2 ^ Q
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
GML 198 1 ^ 9 # DHM < L % A9 :: 5 < ZV 28 TIQU < 2535 % AXG % L 1UM8 :: 2 ^ Q
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
81X9 # 98 > T // GWB :: NOAGWB :: 7 = ? 9WNOAGWB :: E = ? 9WB8FW ? 9B7B1
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
GML 198 1 ^ 9 # DHM < L % A9 :: 5 < ZV 28 TIQU < 2535 % AXG % L 1UM8 :: 2 ^ Q
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
2 ^ Q 08A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
GML 198 1 ^ 9 # DHM < L % A9 :: 5 < ZV 28 TIQU < 2535 % AXG % L 1UM8 :: 2 ^ Q
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
81X9 # 98 > T // GWB :: NOAGWB :: 7 = ? 9WNOAGWB :: E = ? 9WB8FW ? 9B7B1
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
GML 198 1 ^ 9 # DHM < L % A9 :: 5 < ZV 28 TIQU < 2535 % AXG % L 1UM8 :: 2 ^ Q
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
GML 198 1 ^ 9 # DHM < L % A9 :: 5 < ZV 28 TIQU < 535 % AXG % L 1UM8 :: 2 ^ Q
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
81X9 # 98 > T // GWB :: NOAGWB :: 7 = ? 9WNOAGWB :: E = ? 9WB8FW ? 9B7B1
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
GML 198 1 ^ 9 # DHM < L % A9 :: 5 < ZV 28 TIQU < 2535 % AXG % L 1UM8 :: 2 ^ Q
2 ^ Q 18A :: FNOAG // XCKO :: ? 9 BEB 17 B = 6 > ? 9 BEBV7BUZV28TZQU ? <
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::
4HS83Q % AXGX98 1 ^ 9 # 98 1 ^ 98FW ? 9 > ? // GWB :: 1GB = 6 % AXG 198 1 ^
BEB :: % 8NOAGWBE7 = ? 9WNOAGWB ? EC ? 9WB8FW ? 9 BEB 17B98 1% > ?
WW8FW ? 9 BEB 17 OAGWBO7 = ? 9W ^ 8FW8FW ? 9 BEBW8FW8FW ? 9 BEB ::



6. SAMPLING RESULTS

The sampling implementation process was carried out by a cross-organisational team of colleagues from ACER, representatives from the Community of Madrid Education Department, and the British Council in Spain and the UK. With

focus specifically on the comparison unit of the Madrid region of Spain, all students fulfilling the target population criteria detailed in Chapter 3 are described in Table 6.

Table 6: Target population definition in Madrid, Spain

Grade 10 in Madrid region	Years of formal schooling	Average age at time of testing	Information about age of entry, promotion and retention
ESO 4	10	15.6	Children must be six years old to enter primary education. Primary education is free and compulsory for all children. Students who fail the end-of-year exam are held back

The target school population within the region was 100 per cent of all publicly funded schools in Madrid. There were no school exclusions made. A small number of students were classified as having a functional or intellectual disability and therefore exempt from taking

part in the assessment, as described in Table 7. The overall rate of school level and within-school exclusions was within the rate of five per cent outlined in participation standard 1.3. Therefore, a high standard of participation was successfully achieved.

Table 7: Coverage and exclusions

Coverage	Notes on coverage	School-level exclusions	Within-sample exclusions/refusals	Overall exclusions/refusals
100%	All publicly funded schools in Madrid	0.0%	4.5%	4.5%

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 4 1 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

As described in Chapter 4, a process of implicit stratification was used to ensure a proportional sample allocation across all implicit strata. The stratification variables applied to the Madrid sample frame were bilingual education, school type and geographic region.

The first stage of the two-stage cluster sample process was to draw the school sample from the

complete school sample frame. A total of 170 schools were drawn, as shown in Table 8. The sample process involves the selection of two substitute schools that can be used in the event of the first sampled school not being able to participate. None of the substitute schools was used within the Madrid region. There was no participation from one of the original sampled schools due to an administrative complication.

Table 8: School sample size

Number of schools in original sample	Number of eligible schools in original sample	Number of schools in original sample that participated	Number of replacement schools that participated	Total number of schools that participated
170	170	169	0	169

The second stage of the two-stage cluster sampling process was the random selection of eligible students, from the target grade, within each participating school to take part in the assessment. Table 9 shows the total number of

students in all sampled schools, the total number withdrawn, excluded, eligible and absent on the day of the assessment. The total number of students that took part in the assessment across the region was 1,774.

Table 9: Student sample size

Within-school student participation (weighted percentage)	Number of sampled students in participating schools	Number of students withdrawn from school	Number of students excluded	Number of eligible students	Number of students absent	Number of students assessed
90%	2,028	78	14	1,936	162	1,774

A comparison between the school and student population and an estimate calculated from the draw sample is carried out as a check on the accuracy of the sampling procedure, as shown in Table 10. The table shows the actual number of schools and students within the Madrid region and the number of participating sampled schools and students, and an estimate of the student population size based on the sample data. The population figures are derived from

the sampling frame used to select the sample, while the achieved sample figures refer to the number of sampled schools and students that participated in the assessments. The achieved sample figures were calculated using sampling weights and used as a check on the accuracy of the sampling procedure. As shown, the population size estimated from the sample closely matched the population size from the sampling frame.

Table 10: Population and sample size

Madrid population		Achieved sample population		
Schools	Students	Schools	Students	Student population size estimated from sample
707	51,947	169	1,936	51,580

The English Impact Madrid response rate is an important participation standard and indication of the successful implementation of the school and student sampling procedure. Table 11 shows the weighted school and student participation rates. The weight applied to each school corresponds to the number of schools and students they represent in the entire population.

For the case of Madrid, the population was divided into two strata: bilingual and non-bilingual schools. Each student within each school had a weight equal to:

$$\frac{\text{Total population of students in stratum}}{\text{Total number of students participating in assessment}}$$

Table 11: Weighted school and student participation rates

School participation		Student participation	Overall participation	
Before replacement	After replacement		Before replacement	After replacement
99.3%	99.3%	87.3%	86.7%	86.7%

The weight applied can vary from school to school. On average, each participating student in stratum 1 represents 20 students from the total population of students in bilingual schools. However, every student participating from non-bilingual schools represents, on average, 34 students from the whole population of students in non-bilingual schools. The weighted response rates take into account the weight

each school has in the total sample, that is, the number of students it represents. As outlined in participation standard 1.7, the main survey response rate should reach at least 85 per cent of all sampled students across responding schools, with an overall participation rate after replacement of schools of 86.7 per cent. The Madrid region exceeded this standard.

7. ASSESSMENT OUTCOMES

34% of students from the Madrid comparison unit achieved B2 or C level on the CEFR scale

The English language assessment taken by all participating students contained five components including grammar and vocabulary, testing four language learning skills: listening, reading, speaking and writing.

Over one-third of the comparison unit population from Madrid (34 per cent) achieved B2 or C level in their overall English language test performance; 38.5 per cent of the participating population achieved B1 level on the CEFR; 23.4 per cent at A2 level; and 4.2 per cent at A1 level as shown in Table 12.

Table 12: Distribution of student population for overall test performance by CEFR level

CEFR level	Overall English language test performance		
	Percentage (%)	Standard error (%)	95% CI
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
A0	0	-	-
Total	100		

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 4 5 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

As shown in Table 13, the standard error of the mean estimates for all four skills is small, indicating a high level of precision for these estimates.

Table 13: Mean population score and CEFR levels by skills and overall achievement

Overall score	Grammar vocabulary	Listening	Reading	Speaking	Writing	Overall	Weighted cases	Un-weighted cases	Schools
Mean	24.38	35.39	29.44	28.31	30.14	123.28	51,580	1,774	169
CEFR	-	B1	B1	B1	B1	B1			
SE	0.28	0.30	0.40	0.48	0.43	1.48			
95% CI	(23.83, 24.93)	(34.80, 35.99)	(28.65, 30.24)	(27.36, 29.25)	(29.29, 30.98)	(120.37, 126.19)			

Listening

The highest mean scale score achieved – 35.39 per cent

Receptive skills

Participating students' receptive skills, listening and reading were assessed using the computer-based test delivered via tablet using individual headphones for the listening component.

The highest mean scale score was for the listening skill, 35.39 per cent, as shown in Table 13 (see page 41). The estimated distribution of the listening skill performance, shown in Table 14, shows that 44.6 per cent of participating students achieved at B1 level, while a high proportion, 39.9 per cent, achieved at B2 or C level for their listening skill.

Table 14: Estimated distribution of student population for listening and reading skills by CEFR level

CEFR level	Listening component			Reading component		
	Percentage (%)	Standard error (%)	95% CI	Percentage (%)	Standard error (%)	95% CI
C	19.4	1.0	17.4, 21.4	17.3	1.06	[15.2, 19.4]
B2	20.5	1.1	18.4, 22.7	15.5	0.93	[13.7, 17.3]
B1	44.6	1.2	42.1, 47.1	31.2	1.17	[28.9, 33.4]
A2	15.3	1.01	13.3, 17.3	32	1.25	[29.6, 34.5]
A1	0.1	0.06	0.0, 0.2	4	0.62	[2.8, 5.2]
A0	0.1	0.08	0.0, 0.3	0	-	-
Total	100			100		

The mean achievement for the Madrid population in reading skills is 29.44 per cent, the third highest of the four components assessed. As shown in Table 14, 31.2 per cent of participating students achieved at B1 level

on the CEFR. A smaller proportion achieved above B1 level, and 32.8 per cent when looking at the distribution of reading achievement across the population.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 4 7 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Productive skills

The productive skills – speaking and writing – were also tested using the computer-based English language assessment via tablet with individual headphones and microphone to capture speech responses and a keyboard to enable participating students to type with as much ease as possible.

Speaking skills achieved the lowest mean scale score, 28.3 per cent, as shown in Table 13. The distribution of speaking skills when referenced against the CEFR highlights an interesting trend: despite being the lowest mean score across the four language learning skills, the largest proportion of students achieved at B1 level, 47.7 per cent (see Table 15). The lowest proportion of students when comparing the estimated distribution across all four skills achieved at B2 or C level – 17.9 per cent.

Table 15: Estimated distribution of student population for speaking and writing skills by CEFR level

CEFR level	Speaking			Writing		
	Percentage (%)	Standard error (%)	95% CI	Percentage (%)	Standard error (%)	95% CI
C	2.3	0.37	1.6, 3.0	4.7	0.58	3.6, 5.9
B2	15.6	1.12	13.4, 17.8	25.7	1.46	22.9, 28.6
B1	47.7	1.49	44.7, 50.6	35.5	1.25	33.1, 38.0
A2	19.1	1.23	16.7, 21.5	17.5	1.07	15.4, 19.6
A1	10.1	0.93	8.3, 11.9	14.3	0.99	12.3, 16.2
A0	5.2	0.78	3.7, 6.8	2.3	0.51	1.3, 3.3
Total	100			100		

The second highest mean scale score was achieved for writing skills: 31.14 per cent (see Table 13). The estimated distribution of writing performance when referenced against the CEFR shows that 35.5 per cent of all students achieved at B1 level (see Table 15). The estimated distribution shows a high proportion of students, 30.4 per cent, achieved at B2 or C level in writing skills.

Female students consistently achieved higher mean scale scores for all skills than males

Comparing achievement by gender

The ratio of participating female and male students was evenly distributed at 50.9 per cent and 49.1 per cent respectively.

As shown in Table 16, female students consistently achieved higher mean scale scores for all test components, skills and when calculating an overall score. On average, females scored 1.28 points higher than males for each test component and skill.

Table 16: Mean performance score by gender, skills and overall achievement

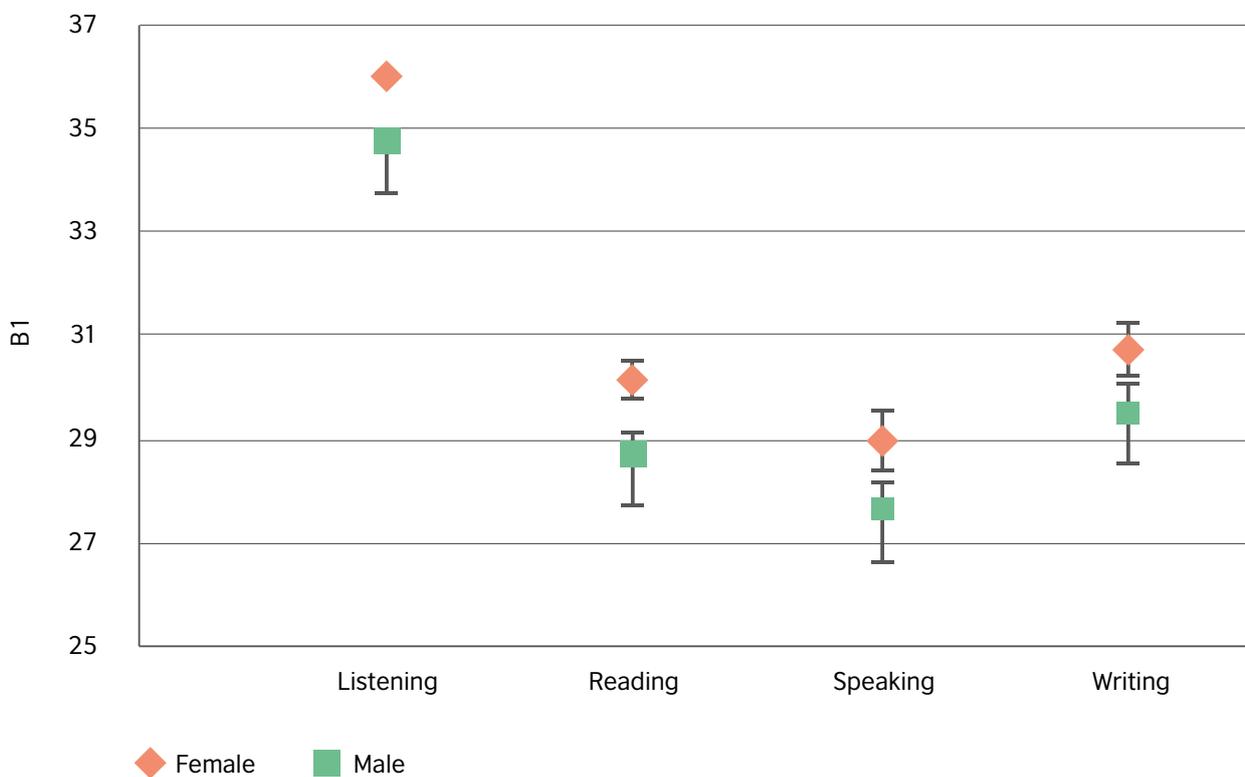
	Grammar/ vocab	Listening	Reading	Speaking	Writing	Overall	Weighted cases	Un- weighted cases	Schools
Female	24.92	36.01	30.15	28.98	30.74	125.88	26,222	903	168
Male	23.82	34.76	28.72	27.61	29.51	120.60	25,358	871	168
Difference (F-M)	1.10	1.25	1.43	1.37	1.23	5.28			
Standard error	0.336	0.369	0.574	0.534	0.493	1.673			
Comparison (95% confidence)	▲	▲	▲	▲	▲	▲			

Key: Positive difference ▲ No difference ► Negative difference ▼

Comparative analysis of the CEFR performance of female and male students, shown in Figure 8 below, shows that the gender performance profile across the four skills remains the same: listening is the strongest skill, followed by writing, reading and speaking. Females and males achieve on average within B1.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 4 9 3 Q J H
G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Figure 8: Comparative mean CEFR performance of females and males by skills and overall achievement



Students from bilingual schools achieve on average 6.9 points higher than students at non-bilingual schools across all four skills

Comparing achievement by bilingual education status

One of the central research aims at the heart of English Impact Madrid was to understand the comparative capability across students participating in the bilingual programme and those in schools not participating in the regional initiative. The active oversampling of bilingual schools, as described in Chapter 4, was designed to allow for this comparison.

The comparative performance of students at bilingual and non-bilingual schools showed a considerable difference. Students from bilingual schools on average achieve 6.9 scale score points higher than students at schools not participating in a bilingual programme.

Table 17: Mean performance score by bilingual and non-bilingual schools – skills and overall achievement

	Grammar/vocab	Listening	Reading	Speaking	Writing	Overall	Weighted cases	Un-weighted cases	Schools
Bilingual	29.19	40.09	35.33	34.52	35.95	145.88	10,514	543	50
Non-bilingual	23.15	34.19	27.94	26.72	28.65	117.50	41,066	1,231	119
Difference (B-NB)	6.04	5.90	7.39	7.80	7.30	28.38			
Standard error	0.599	0.526	0.752	0.855	0.800	2.678			
Comparison (95% confidence)	▲	▲	▲	▲	▲	▲			

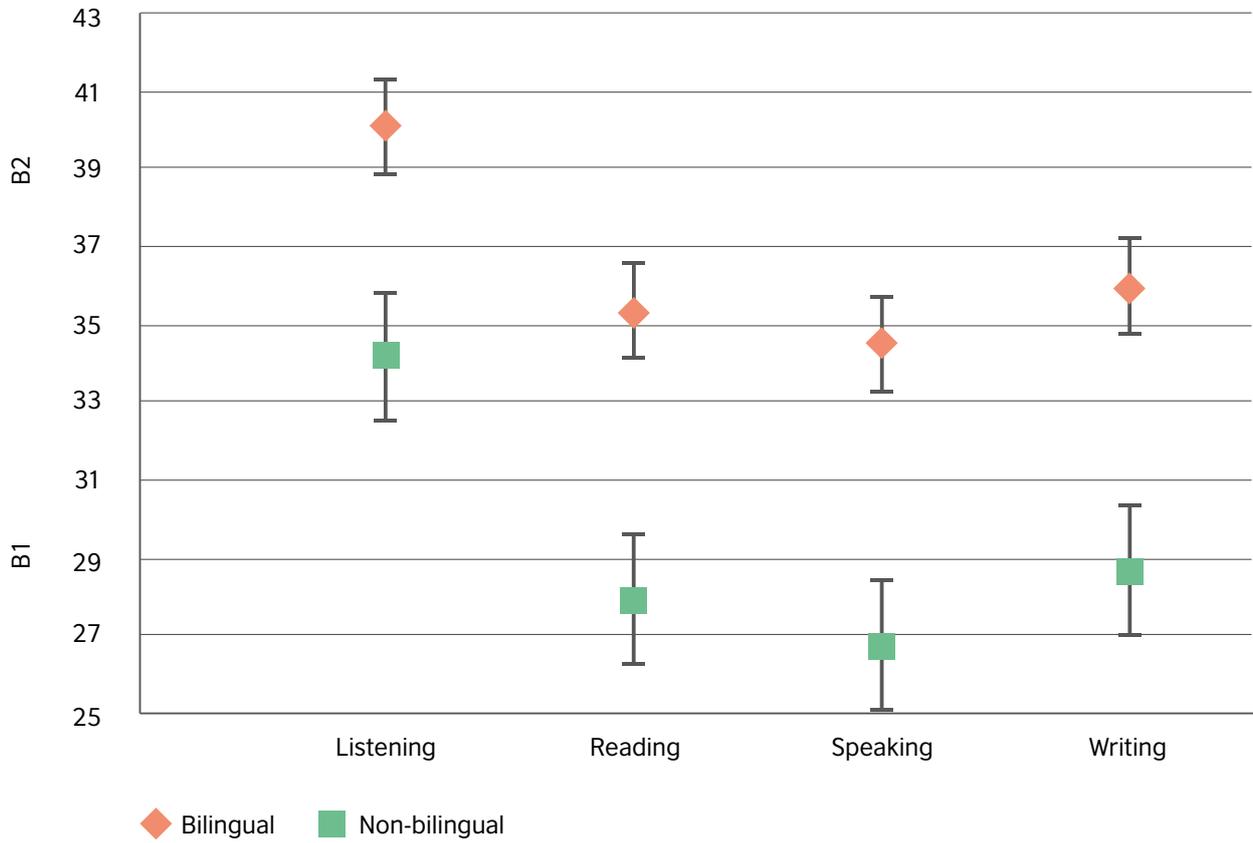
Key: Positive difference ▲ No difference ► Negative difference ▼

The language learning skills profile of students at bilingual and non-bilingual schools mirrored that of the total Madrid population: listening was assessed as the strongest skill, followed by writing, reading and speaking. However, when looking at the comparative mean CEFR

performance of students from both types of schools, students participating in the bilingual programme achieved on average at B2 level for their listening skills when referenced to the CEFR.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 5 1 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Figure 9: Comparative mean CEFR performance of students at bilingual and non-bilingual schools by skills and overall achievement



Across the Madrid region there are two streams of bilingual schools: *sección*, where students study three or more subjects in English, and *programa*, in which pupils study one or two subjects in English. To gauge the performance of students participating in these two streams,

responses to the following questionnaire item were grouped to create the corresponding variables: how many subjects, other than in dedicated English language lessons, do you learn in English?

Table 18: Count and percentage of students participating in programa and sección bilingual streams by overall CEFR performance

Bilingual stream		Overall CEFR level				
		A1	A2	B1	B2	C
Programa	Count	5	53	124	62	42
	% within programa	1.7%	18.5%	43.4%	21.7%	14.7%
Sección	Count	0	4	28	105	101
	% within sección	0%	1.7%	11.8%	44.1%	42.4%

The programa and sección streams within the bilingual programme were not included in the original stratification process, and therefore the above figures are shown as basic counts of participating students and percentages to which weighting has not been applied. These figures do provide some preliminary indication that proficiency levels are different between these groups. For example, 86.5 per cent of students participating in the sección stream, receiving more subject lessons taught in English, are achieving significantly at B2 and C level. These preliminary findings suggest that further targeted research is required to assess whether these differences are real and significant for the entire bilingual student population.

85%

of students participating in the sección stream achieve B2 and C CEFR level

Comparing achievement by charter and state school populations

As detailed in Table 4 (see page 27), the implicit stratification variables included in the sample design included the reference of school type as charter or state school. State schools are described as non-denominational schools fully funded by the regional government; charter schools are described as private schools receiving a regional government subsidy to provide education services.

Table 19 shows the mean performance score for both charter and state schools for each test component and language learning skill, also showing the overall scale score performance of both groups. Charter schools consistently achieve higher mean scale scores for all skills, on average scoring 2.8 scale score points above state schools. Most notably, the mean performance estimates achieved for speaking skills across the student population from charter schools was 4.63 scale score points higher than the estimates for state schools.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 5 3 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Table 19: Mean performance score by charter and state schools – skills and overall achievement

	Grammar/ vocab	Listening	Reading	Speaking	Writing	Overall	Weighted cases	Un- weighted cases	Schools
Charter schools	25.33	36.54	31.01	30.98	31.91	130.45	21,741	673	63
State schools	23.69	34.56	28.30	26.36	28.85	118.06	29,839	1,101	106
Difference (C-S)	1.64	1.98	2.71	4.63	3.07	12.38			
Standard error	0.607	0.665	0.873	0.993	0.915	3.186			
Comparison (95% confidence)	▲	▲	▲	▲	▲	▲			

Key: Positive difference ▲ No difference ► Negative difference ▼

In order to more fully understand the performance of students at charter and state schools, the analysis in Table 20 was carried out showing the count and percentage of students from both types of schools' overall CEFR performance.

Table 20: Count and percentage of students from charter and state non-bilingual schools by overall CEFR performance

Type of school		Overall CEFR level				
		A1	A2	B1	B2	C
State	Count	40	218	232	66	21
	% within state schools	6.9%	37.8%	40.2%	11.4%	3.6%
Charter	Count	21	112	275	194	71
	% within charter schools	3.1%	16.6%	40.9%	28.8%	10.5%

Similarly to the analysis in Table 18, the sample weighting has not been applied to these figures and therefore the precision of these estimates is not reported. However, some preliminary indication can be drawn from the distribution of performance at state and charter schools.

Additional comparison of students from charter schools (N = 673), who are not currently participating in the bilingual programme, and state bilingual (N = 524) and non-bilingual (577) schools, comparing their overall CEFR performance shows that the largest number

of students from state schools participating in the bilingual programme achieve at a much higher level. In real terms 143 students, or 8.1 per cent of the overall total number of participants achieving at C level on the CEFR.

Table 21: Count and percentage of students from charter and state bilingual and non-bilingual schools

Overall CEFR	A1	A2	B1	B2	C	Total participants
School type						
State Bilingual	5	57	152	167	143	N = 524
% of total	0.3	3.2	8.6	9.4	8.1	
State Non-bilingual	40	218	232	66	21	N = 577
% of total	2.3	12.3	13.1	3.7	1.2	
Charter Non-Bilingual	21	112	275	194	71	N = 673
% of total	1.2	6.3	15.5	10.9	4	

Comparing achievement in geographic regions of Madrid

The third and final implicit stratification variable was administrative region, allowing a level of analysis of school populations, comparative performance in the capital, and east, north, west and south of the Madrid region.

As shown in Table 21, the western region has on average a higher mean performance than all other regions. The southern region underperforms with respect to the capital and western regions, but has a similar performance to those of the north and east.

Table 22: Pairwise comparison of mean performance of administrative regions of Madrid

	Overall score (SE)	Capital	East	North	West	South
Capital	126.0 (2.48)	-	0	0	q	p
East	120.2 (4.05)	0	-	0	q	0
North	120.3 (3.58)	0	0	-	q	0
West	135.3 (3.84)	p	p	p	-	p
South	115.5 (2.80)	q	0	0	q	-

- p Region on the left higher than region at the top
- q Region on the left lower than region at the top
- 0 No difference between the two regions

8. ENGLISH LANGUAGE LEARNING MOTIVATION

The analysis presented in this chapter explores two key areas:

1. the levels and character of language learning motivation for different groups of students (split by gender, bilingual and non-bilingual schools and socio-economic status (SES))
2. relationships between each of the motivational scales and English language proficiency.

The student context questionnaire received a full response rate, largely owing to the fact that students could not proceed through the app without answering all the questions. However, a good range of responses was recorded which indicates that participants took the survey seriously and gave it their attention and consideration. This is important, since the analysis described in the upcoming section accounts for nuances in the patterns of response to this part of the survey.

CONSTRUCT VALIDATION FINDINGS

The measurement model

As described in Chapter 3, the aim of the construct validation exercise was to establish a measurement model that satisfactorily reflects the pattern of observed responses. The value of a well-fitting measurement model is that it provides a statistically sound means of estimating levels of motivation on each of the scales for all participating students.

This, in turn, enables comparisons to be drawn between groups of students using further CFA modelling techniques.

The hypothesised measurement model has the structure shown in Figure 10 (see page 54). Please cross-reference terms with Table 2. (see page 17) This initial model showed this structure to have a reasonable fit to the data according to accepted thresholds for CFA models (CFI = 0.930; TLI = 0.920; RMSEA = 0.051). It was found that the model could be improved slightly by allowing some of the error terms (or residual error) for individual questions to correlate with each other. These relationships are indicated in red on Figure 10.⁴ In the current modelling exercise, the inclusion of four covariances between error terms was found to make a substantial improvement to the model. These are listed in Table 23. The error covariances were added in order of their estimated impact on model fit using calculations of adjusted model chi-square (Satorra and Bentler, 2010), and while further covariances could be conceivably included, they were found to have minimal impact on the adjusted chi-square values and overall model fit.

4. Decisions regarding the inclusion of the covariances mentioned in Table 22 were made with reference to the modification indices generated by the Mplus software. However, no covariances were included that did not make sense substantively. Hence, these were restricted to covariances between errors for questions against the same motivational scale.

Table 23: Improvements in adjusted chi-square values following addition of covariances between error terms

Covariance	Improvement in adjusted X ² on 1 degree of freedom	P-value
INSTRU2 with INSTRU1	197.127	<.001
IO4 with IO3	83.636	<.001
O4 with O1	53.070	<.001
EX3 with EX1	40.904	<.001

Including these enables the model to take into account commonality between two observed variables in addition to that explained with reference to the latent variable, perhaps reflecting something in the wording of both questions that provokes a particular shared response.⁵ The final measurement model therefore included the four correlated error variances listed in Table 23. The model provides a good fit for the data (CFI = 0.945; TLI = 0.937; RMSEA = 0.045). Final model estimates are given in Appendix B.

5. See Brown (2015, pages 157–162) for a fuller discussion of correlated measurement error and the implications in CFA.

A :: FNOAG // XCKPW ? 9BE // 17B = 6 > ? 9BEB 17B = ZV 28TZQU - < 2535 % A
 NOAGWBN7 = ? 9WNOAGWB :: E = ? 9WB8FW ? 9BEB 17B981% > ? 9BEB 17 ?
 ? // GW 5816B ENGLISH ? IMPACT B :: 17B = 6PW8FW ? 4 > JHY4HS83QJH
 G % 2AXGML 1981 ^ 9 # DRM < L % A9M5 < ZV 28TIQU < 2535 % AXG % L 1UM8
 FWB :: 17B = 6PW ? 9BEB 17OAGWB :: 7 = ? 9 = W8FW8FW ? 9BEBW8FW8FW ?

Figure 10: Measurement model

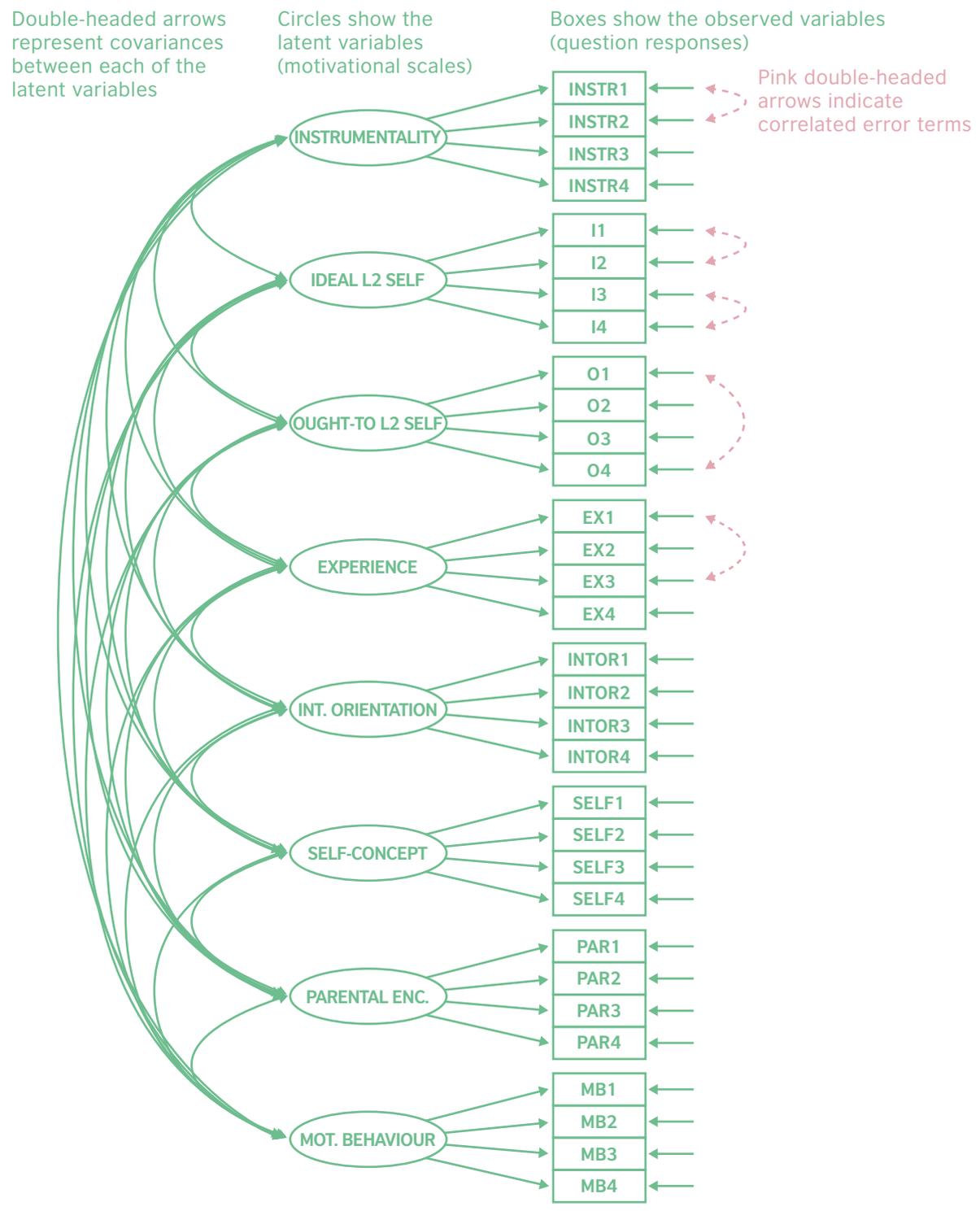


Table 24: Error covariances included in the measurement model

Covariance ref.	Question ref.	Question details	Reduction in adjusted chi-square value	Significance
1	INSTR1	I need English for my future career.	163.005	p<.001
	INSTR2	The things I want to do in the future require me to use English.		
2	I3	In the future, I would really like to communicate with people from other countries online.	90.492	p<.001
	I4	If I could speak English well, I could get to know more people from other countries via the internet.		
3	EX1	Learning English is really great.	47.477	p<.001
	EX3	I find learning English really interesting.		
4	O1	I consider learning English important because the people I respect think that I should do it.	43.169	p<.001
	O4	Learning English is necessary because people surrounding me expect me to do so.		

The structural model

Having established the viability of the measurement model, it is insightful to examine the association between the latent variables as estimated by the model. Standardised covariances (correlations) estimated by the model are given in Table 23 (see page 57). The values given range from those classified as low to strong correlations (Cohen, 1988), suggesting that the strength of relationships between variables differs.

Two latent variables in particular are clearly less related to other areas of motivation. These are parental encouragement and the ought-to L2 self. This could be ascribed to the fact that, unlike other latent variables, they focus on external pressures connected with studying English. The first one refers to levels of parental

encouragement and the second one relates to pressures from the broader environment by representing what is expected of the participants in terms of studying English. In this sense, they can be seen as more externalised motives to studying English (Ryan and Deci, 2000).

Most importantly, both parental encouragement and the ought-to L2 self scales are least closely related to motivated learning behaviour, the latent variable that aims to capture reported levels of effort invested in language learning. This finding is in line with previously reported results, which indicate that the influence of parental encouragement and the ought-to L2 self on motivated behaviour is limited (Csizér and Kormos, 2009; Iwaniec and Ullakonoja, 2016; Taguchi et al. 2009).

Table 25: Standardised (STDYX Standardisation) correlations between variables in the final model

Latent variables (short ref.)		Two-tailed			
		Estimate	SE	Est./SE	P-value
EXPER	INTER	0.673	0.020	34.042	0.000
IDEAL	EXPER	0.694	0.016	42.484	0.000
IDEAL	INTER	0.706	0.021	33.552	0.000
INSTR	EXPER	0.563	0.025	22.361	0.000
INSTR	IDEAL	0.623	0.024	25.928	0.000
INSTR	INTER	0.712	0.025	28.055	0.000
MOTIV	INTER	0.468	0.023	20.552	0.000
MOTIV	IDEAL	0.478	0.021	22.421	0.000
MOTIV	INSTR	0.531	0.024	22.057	0.000
MOTIV	EXPER	0.735	0.016	44.839	0.000
OUGHT	IDEAL	0.206	0.026	7.811	0.000
OUGHT	EXPER	0.258	0.027	9.479	0.000
OUGHT	MOTIV	0.280	0.027	10.526	0.000
OUGHT	INTER	0.324	0.027	11.936	0.000
OUGHT	INSTR	0.504	0.028	18.204	0.000
PAREN	EXPER	0.271	0.027	10.055	0.000
PAREN	MOTIV	0.310	0.025	12.191	0.000
PAREN	IDEAL	0.351	0.025	13.803	0.000
PAREN	OUGHT	0.468	0.025	18.514	0.000
PAREN	INTER	0.474	0.028	17.025	0.000
PAREN	INSTR	0.687	0.026	26.635	0.000
SELF	OUGHT	0.189	0.026	7.257	0.000
SELF	PAREN	0.255	0.025	10.203	0.000
SELF	INTER	0.451	0.024	18.641	0.000
SELF	MOTIV	0.476	0.022	21.601	0.000
SELF	INSTR	0.509	0.025	20.499	0.000
SELF	EXPER	0.666	0.017	38.714	0.000
SELF	IDEAL	0.724	0.015	49.657	0.000

Multi-group analysis findings

Comparative analyses were conducted for three groupings of the data:

- gender
- schooling type (bilingual/non-bilingual)
- socio-economic status (SES).

The final sample included 1,774 students and was well balanced in terms of gender with 903 female and 870 male respondents. In addition, 524 students were enrolled in bilingual schools (263 females and 261 males) and 1,249 (640 females and 609 males) were enrolled in non-bilingual schools. Based on the calculations described in the later sections of this report, the students were categorised into two SES groups: lower SES (886 students) and higher SES (887 students).

Separation between the groups was assessed, in particular with regard to schooling type and SES group. Bilingual schools are intended to be available across the socio-economic spectrum, and the information derived from the data supports this to some extent. Chi-square analysis indicated although there is a slight weighting towards students in the higher SES group attending bilingual schools ($X^2 = 4.274$ on 1 DF, $p=.039$), the distinctions between membership are enough to justify looking at each group separately. Table 26 gives the breakdown of students in each SES group attending bilingual/non-bilingual schools.

Table 26: Breakdown of students at bilingual/non-bilingual schools by SES group

SES group	Schooling	
	Bilingual	Non-bilingual
Lower	242 (46.2%)	644 (51.6%)
Higher	282 (53.8%)	605 (48.4%)
Total	524 (100%)	1249 (100%)

FEMALE AND MALE STUDENT POPULATIONS' MOTIVATION

The first group comparison is between female and male students. It is prudent not to assume identical motivations drive these groups since empirical studies consistently suggest that female students tend to be more motivated to study English than their male peers (see Iwaniec (2015) for a review). Studies point to higher achievement in language learning among female learners than male learners (Fernandez Fontecha, 2010; Jimenez Catalan, 2010). Considering that motivation is shown to affect language learning achievement (Hsieh and Kang, 2010; Kim and Kim, 2014; Marsh and Martin, 2011), understanding where the differences

lie with respect to underlying motivational traits will help formulate targeted policies on foreign language learning and teaching.

Levels of motivation

Significant differences across genders were found between the levels of motivation reported on six out of eight motivational scales. For five of these, female students reported higher levels of motivation than their male peers (significance level from *t*-test indicated in parenthesis):

- motivated behaviour ($p < .001$)
- international orientation ($p < .001$)
- ideal L2 self ($p = .018$)
- language learning experience ($p < .001$)
- instrumentality ($p = .015$).

Table 27: Breakdown of students at bilingual/non-bilingual schools by SES group

Motivational variable	Group comparisons				t-test statistics		
	gender	N	Mean	Std. Deviation	t	Degrees of freedom	Sig. (two-tailed)
International orientation	female	903	.03306	.335924	4.470	1771	.000
	male	870	-.03855	.338546			
Ideal L2 self	female	903	.04621	.902136	2.374	1771	.018
	male	870	-.05210	.838797			
Parental encouragement	female	903	.02099	.617192	1.207	1771	.227
	male	870	-.01402	.603217			
English self concept	female	903	.03812	1.051489	1.644	1771	.100
	male	870	-.04211	1.001915			
Language learning experience	female	903	.12767	.826413	6.922	1771	.000
	male	870	-.14172	.811605			
Instrumentality	female	903	.03796	.633502	2.435	1771	.015
	male	870	-.03640	.652269			
Motivated behaviour	female	903	.19608	.874410	9.193	1771	.000
	male	870	-.18369	.864653			
Ought-to L2 self	female	903	-.03003	.676285	-1.971	1771	.049
	male	870	.03594	.732774			

The results imply that, compared to male students, female learners reported:

- investing more effort in language learning
- to be more strongly oriented to learn English for the purpose of communication with other speakers of this global language
- to have more robust visions of themselves as successful English users in the future
- to have a more positive language learning experience, conceptualised as language learning attitudes
- to be more likely to learn English with a view to using it to progress in their future careers.

These results are consistent with findings from previous empirical studies (see Iwaniec (2015) for a review). Previous research in the area implies that there might be a number of reasons why female learners express higher motivation than male learners.

- General perceptions of language learning as a female domain foster the construction of femininity among females learning languages, whereas they have the opposite effect for the perceptions of masculinity of males learning a foreign language (Carr and Pauwels, 2005b).
- Language proficiency is perceived as more directly relevant for the future careers of female than male learners (Clark, 1998) – a perception that is even more augmented by the perceived limited choice of professional careers for females (Norton and Pavlenko, 2004).
- Females are perceived as having a stronger preference for, and value communication more, than male learners (Chávez, 2000).

- Preferred learning styles and use of language learning strategies of female learners (visual and auditory learning) is more convergent with the requirements of language classes than in the case of male learners who are often kinaesthetic learners (Oxford, 1993).

In the case of one motivational trait – the ought-to L2 self ($p=.049$) – male learners' score was higher than that of female learners. As the p value is very close to the .05 threshold, extra caution needs to be taken when considering this result. The finding suggests, however, that male learners are more likely than female learners to perceive English proficiency as a requirement imposed on them by society at large.

No statistically significant gender difference was found for two motivational scales: English self-concept and parental encouragement. The latter finding indicates that both male and female students reported to be equally supported and encouraged by their parents to study English.

With respect to English self-concept, this can be interpreted as indicating something about the confidence of female learners in their ability to learn English. Despite multiple reasons for higher language learning motivation of female than male learners, Spanish female learners did not report themselves to be better at learning English than their male peers. This is consistent with previous findings, which imply that females tend to be more conservative when estimating their abilities than male learners (Bandura, 1997; Edele et al., 2015; Skaalvik and Skaalvik, 2004).

Relationships between motivational scales

Analysis shows that freeing up three covariances leads to a significant improvement in the adjusted model chi-square (Satorra and Bentler, 2010). This means that these relationships are significantly different between female and male students:

- parental encouragement with ideal L2 self (adjusted X^2 : 10.658 on 1 DF, $p < .001$)

- language learning experience with international orientation (adjusted X^2 : 10.527 on 1 DF, $p < .001$)
- instrumentality with Ideal L2 self (adjusted X^2 : 3.925 on 1 DF, $p = .0475$)

This results in a model with good overall fit statistics (CFI = 0.938; TLI = 0.935; RMSEA = 0.046). The covariance estimates for this gender-specific model are given in Table 27.

Table 28: Breakdown of students at bilingual/non-bilingual schools by SES group

Covariance (using short ref.)	Estimate	SE	Est./SE	Two-tailed p -value
Paren. with ideal				
Female model	0.280	0.030	9.363	0.000
Male model	0.433	0.028	15.273	0.000
Exper. with inter.				
Female model	0.639	0.024	26.242	0.000
Male model	0.707	0.022	31.566	0.000
Instr. with ideal				
Female model	0.581	0.028	20.436	0.000
Male model	0.671	0.028	23.686	0.000

In all cases of covariances that were found to be significantly different, the strength of the covariance is higher for male than female learners. The findings imply that:

- parental encouragement is more strongly related to creating a vision of oneself as a successful language learner for male than female learners
- positive language learning experience is more closely related to perceptions of English as a lingua franca for male than female learners
- male learners' visions of themselves as successful language learners are more likely to be related to their future professional plans than for female learners.

Overall, these findings tend to suggest that motivation of male learners is more dependent on the external environment (parental encouragement, positive language learning experiences, and the expectations of future career) than is the case of female language learners.

BILINGUAL AND NON-BILINGUAL SCHOOLS' STUDENT POPULATIONS' MOTIVATION

The second group comparison is drawn between students enrolled in bilingual and non-bilingual schools. In bilingual schools, students do not study English solely as a school subject but also study through the medium of English, whereas in non-bilingual schools, studying English is only a subject. Thus, there is a substantial difference between students' language learning environments and their use of and engagement with the English language, which might lead to differences in both language learning motivation and ultimate achievement.

Levels of motivation

Significant differences between the reported levels of motivation for learners from bilingual and non-bilingual schools were found on seven out of eight scales.

The scale for which there was no statistically significant difference found between bilingual and non-bilingual learners was that of instrumentality ($p=.069$). This indicates that learners from both groups report that English will be equally important for their future careers. Lack of significant difference on this scale may reflect the perception that competence in English is increasingly perceived as a basic skill (Erling and Seargeant, 2013), without which young people are not able to compete on the job market (Graddol, 2006).

For five motivational scales, students from bilingual schools reported higher levels of motivation (significance level from t -test indicated in parenthesis):

- international orientation ($p=.001$)
- ideal L2 self ($p<.001$)
- English self-concept ($p<.001$)
- language learning experience ($p<.001$)
- motivated learning behaviour ($p=.018$).

Table 29: T-tests to check for differences between mean-factor scores for bilingual and non-bilingual school groups

Motivational variable	Group comparisons				t-test statistics		
	Schooling	N	Mean	Standard deviation	t	Degrees of freedom	Sig. (two-tailed)
Motivated learning behaviour	Bilingual	524	.0719	.81470	2.250	1771	.025
	Non-bilingual	1249	-.0321	.91761			
International orientation	Bilingual	524	.0383	.29077	2.960	1771	.003
	Non-bilingual	1249	-.0133	.35227			
Ideal L2 self	Bilingual	524	.1982	.78953	6.367	1771	.000
	Non-bilingual	1249	-.0878	.89175			
Parental orientation	Bilingual	524	-.0532	.62902	-2.252	1771	.024
	Non-bilingual	1249	.0179	.59722			
English self-concept	Bilingual	524	.2201	.93283	5.983	1771	.000
	Non-bilingual	1249	-.0966	1.04994			
Language learning experience	Bilingual	524	.1054	.78905	3.525	1771	.000
	Non-bilingual	1249	-.0460	.83984			
Instrumentality	Bilingual	524	.0435	.61478	1.819	1771	.069
	Non-bilingual	1249	-.0174	.65471			
Ought-to L2 self	Bilingual	524	-.0890	.71637	-3.395	1771	.001
	Non-bilingual	1249	.0347	.69307			

This implies that, compared to their non-bilingual peers, students from bilingual schools reported:

- to be more strongly oriented to learn English for the purpose of communication with other speakers of this global language
- to have more robust visions of themselves as successful English users in the future
- to be more confident at learning English
- to have more positive language learning experience, conceptualised as language learning attitudes
- to invest more effort in language learning.

These findings could be explained by reference to differences in language learning environment. As students in bilingual schools are already using English as a means for communication, they are more likely to perceive the benefits of English proficiency and to have internalised English proficiency as part of their future selves than students from non-bilingual schools, for whom English is another academic subject to study. Similarly, the everyday life of bilingual students may provide them with positive feedback on their learning abilities as they are able to follow classes and respond to teachers in English, whereas learners from non-bilingual schools lack these opportunities.

Students from non-bilingual schools reported higher levels of motivation than their bilingual peers on two scales:

- parental encouragement ($p=.024$)
- Ought-to L2 self ($p<.001$).

These findings may imply that learners from non-bilingual schools perceive themselves as receiving more parental encouragement and feeling more pressure from those around them to study English than learners from bilingual schools. This may indicate that compared to students from bilingual school, students from non-bilingual schools have internalised the need to study English to a lesser extent and may rely more on the stimuli from their external environment to motivate them (Ryan and Deci, 2000).

Relationships between motivational scales

The analysis shows that freeing up one covariance leads to a significant improvement in the adjusted model chi-square (Satorra and Bentler, 2010):

- Parental encouragement with international orientation (adjusted X^2 : 12.43 on 1 DF, $p < .01$).

- The covariance estimates for this school-specific model are given in Table 30.

Table 30: Latent covariances allowed to vary between school groups (standardised estimates)

Covariance (using short ref.)	Estimate	Standard error	Estimate/standard error	Two-tailed p -value
Paren. with ideal.				
Bilingual model	0.403	0.039	10.307	0.000
Non-bilingual model	0.503	0.030	16.760	0.000

The covariance between parental encouragement and international orientation is significantly stronger for the non-bilingual group than the bilingual group. This means that levels of parental encouragement are more closely related to students' perceptions of English as a lingua franca if they attend non-bilingual schools than if they attend a bilingual school. While this finding has never before been reported in scholarly literature, a plausible explanation behind this difference might be that learners in bilingual schools experience communication in English on a day-to-day basis, giving them immediate evidence for the role of English for communicative purpose. This direct experience may well supersede parental input in this respect.

SOCIO-ECONOMIC STATUS (SES)

The final group comparison is made between students differing in socio-economic status (SES). Research suggests that educational attainment tends to differ according to students' SES, with students with lower SES attaining less than their peers with higher SES (OECD, 2013). Studies of language learning motivation report similar findings, namely, levels of motivation tend to fall in line with students' SES (Iwaniec, 2015; Kormos and Kiddle, 2013; Lamb, 2012, 2013). Hence, it is vital to examine where the differences on language learning motivation of students from different socio-economic backgrounds lie.

This analysis is based on a measure of SES constructed from the information given by the students to the questions listed in Table 31.⁶

6. The decision was taken not to use the information describing the actual job held by the parent in the analysis. It was noted that certain categories were over-represented (general labour, for example), whereas others were under-represented. This could be due to the age of participants who might tend towards selecting a more generic job description. Instead, the focus was placed on the information regarding the general level of job.

Table 31: Questions used in the construction of SES variable

What is the highest level of school completed by either of your parents? Who completed the highest level of school?	<ul style="list-style-type: none"> - Pre-primary[1] - Primary [2] - Lower secondary [3] - Upper secondary [4] - Post-secondary non-tertiary [5] - Undergraduate [6] - Postgraduate [7]
What level is that job [†] ?	<ul style="list-style-type: none"> - Lower-level support worker [1] - Mid-level skilled worker [2] - Mid-level professional [3] - Senior-level manager [4]
Do you have any of these in your home?	[Grid answer format YES or NO] <ul style="list-style-type: none"> - A tablet computer [1] - A car [2] - A bedroom of my own [3] - Air conditioning [4] - A smart TV [5] - My own smartphone [6]

[†] This refers to the job given in response to Q49 ('What kind of job does the same parent (as Q48) do?')

SES scores are calculated using a data-reduction technique called principal component analysis (PCA) in which each of the relevant variables is used to create a single, weighted scale (after Caro and Cortés, 2012). From this it is possible to calculate each participants' SES score. The weights accorded to each variable are determined by the data, as PCA takes into account the common variation of the variables determined to be of value in describing SES. Based on the inspection of frequencies and distribution of these SES scores, the decision was made to split the sample into two groups, later referred to as students with lower and higher SES. It needs to be noted that as the distribution of data was normal, the difference in the scores on the continuous variable between the students who are in the top of

the lower SES category and the bottom of the higher SES category is, in some cases, minimal. Thus, care needs to be taken when analysing the results. However, for the ease of the process, it was decided that the two-way split was a workable solution.

Additionally, it should be noted that this approach means that no absolute measure of SES was employed. This is considered to be an advantage, as the approach allows development of a measure that most accurately captures SES in a given context. The description of the technique as formative by Caro and Cortés (2012, page 17) relates to the fact that the variables that contribute towards the scale are fixed in the analysis, rather than beforehand.

Levels of motivation

Significant differences were found on all latent variables. In all cases, students with higher SES scores score higher than learners with a lower SES (significance level from *t*-test indicated in parenthesis):

- international orientation ($p < .001$)
- ideal L2 self ($p < .001$)
- English self-concept ($p < .001$)
- language-learning experience ($p = .002$)
- motivated learning behaviour ($p = .007$)
- parental encouragement $p < .001$)
- ought-to L2 self $p < .001$)
- instrumentality $p < .001$.

Table 32: *t*-tests to check for differences between mean-factor scores for SES groups

Motivational variable	Group comparisons				<i>t</i> -test statistics		
	SES	N	Mean	Std. deviation	<i>t</i>	Degrees of freedom	Sig. (two-tailed)
Motivated learning behaviour	Lower	886	-.0577	.91287	-2.683	1771	.007
	Higher	887	.0557	.86598			
International orientation	Lower	886	-.0332	.35338	-4.214	1771	.000
	Higher	887	.0344	.32135			
Ideal L2 self	Lower	886	-.1503	.89370	-7.207	1771	.000
	Higher	887	.1445	.82775			
Parental orientation	Lower	886	-.1048	.64651	-6.830	1771	.000
	Higher	887	.0909	.55681			
English self-concept	Lower	886	-.2247	1.03572	-8.829	1771	.000
	Higher	887	.1975	.97717			
Language learning experience	Lower	886	-.0647	.81130	-3.144	1771	.002
	Higher	887	.0587	.84070			
Instrumentality	Lower	886	-.1081	.65146	-7.136	1771	.000
	Higher	887	.1076	.62069			
Ought-to L2 self	Lower	886	-.0837	.71336	-4.916	1771	.000
	Higher	887	.0796	.68418			

Compared to learners with lower SES therefore, students with higher SES reported:

- to be more strongly oriented to learn English for the purpose of communication with other speakers of this global language
- to have more robust visions of themselves as successful English users in the future
- to be more confidence in their ability to study English
- to have more positive language learning experience, conceptualised as language learning attitudes
- to invest more effort in language learning
- to receive more encouragement and support from their parents to study English
- to feel more pressure that they should know English
- to be more likely to learn English with a view to use it to progress in their future professional career.

These uniform findings clearly point to an advantage for students with a higher SES over students with a lower SES in terms of language learning motivation. The members of the higher SES group are not only more likely to have language learning goals involving English but also to invest more effort in language learning. They also collectively evaluate themselves as more capable of learning English and have more positive language learning attitudes. These findings are in line with previous studies in the field of language learning motivation (Iwaniec, 2015; Kormos and Kiddle, 2013; Lamb, 2012). With higher levels of language learning

motivation, students with a higher SES are likely to achieve higher competence in English than learners with a lower SES that in the future may lead to recreating, or even augmenting, the existing inequalities between the two groups.

Previous research shows that there might a number of explanations why students with a higher SES report higher levels of language learning motivation than their peers with a lower SES.

- Students with a higher SES may have easier access to positive role models. Their parents tend to be better educated, with higher levels of English competence and in better employment, where English is more likely to be used. This vicarious experience strengthens these students' confidence in their ability to learn English and shows them that English is useful on the job market (see Iwaniec, 2015 for a review).
- Higher economic status can enable students with a higher SES to travel abroad more freely, to have access to private education, extra classes and language learning materials that their peers with a lower SES are less likely to enjoy.
- University-educated parents tend to have higher aspirations for their children, engage more in activities promoting learning (Davis-Kean, 2005; Eccles, 1994) and provide a more supportive learning environment for their children (Klebanov et al., 1994) than parents who have lower levels of education.

Relationships between motivational scales

The analysis shows that freeing up one covariance leads to a significant improvement in the SES-adjusted model chi-square (Satorra and Bentler, 2010):

- motivated behaviour with international orientation (adjusted X^2 : 5.844 on 1 DF, $p=.002$).

The covariance estimates for this gender-specific model are given in Table 33.

Table 33: Latent covariances allowed to vary between SES groups (standardised estimates)

Covariance (using short ref.)	Estimate	Standard error	Estimate/standard error	Two-tailed p -value
Motiv. with inter.				
SES lower group model	0.405	0.027	15.086	0.000
SES higher group model	0.543	0.029	18.800	0.000

The relationship between motivated learning behaviour and international orientation is much stronger for higher SES students than for those in the lower SES. This implies that there is a closer link for higher than lower SES students between effort investment in language learning and orientation to learn English for the purpose of future communication. While both groups

consider English an important tool enabling communication, this perception is more likely to be translated into language learning effort for members of the higher SES group. It can be inferred, perhaps, that members of the lower SES group do not associate this communicative goal as something that is as relevant to them.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 7 3 3 Q J H
G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

LINKS BETWEEN MOTIVATION AND PROFICIENCY

In this section, the relationship between Spanish students' language learning motivation and their proficiency in English is examined. While this is clearly of interest at policy level, there are important caveats to interpreting the findings of such analysis in the current context. First, the reader needs to keep in mind that language learning can be affected by a plethora of factors (Ortega, 2009), only one of which is motivation. Some of them are specific to individuals. For example, aptitude or gift for languages, anxiety, language-learning strategies, cognition and personality traits. Others are specific to the language studied, for example the extent to which it is similar to/different from learners' first language, or the role this language plays in a given context. The amount of exposure to language and the opportunities to use this language are also crucial. In addition, it should be noted that in this study, language-learning motivation of students from Madrid was measured at a time they took the proficiency test. However, motivation is dynamic and changes over time. Hence, whereas the proficiency measure is a cumulative measure of what students have achieved over years of learning, the data on motivation presents just a single snapshot. Considering these points, it is expected that motivation can explain only some variance in proficiency. Nonetheless, a focus on the comparative values leads to some interesting insights.

Presentation of the results looks first to give an overview of the motivational profiles for students who achieve at different levels on the overall CEFR scale, followed by correlations between Aptis test scores and motivational scales for gender, school and SES group. Implications of the findings are discussed.

Learner proficiency and motivational profiles

This analysis groups the students by their CEFR level and compares the motivation profiles for each group based on the calculation of factor scores (F-scores) derived from the CFA described above. As the number of students who achieved the A1 level of proficiency was low, this group was merged with those who achieved an A2 level. Overall, four groups are compared: A (A1 and A2), B1, B2 and C.

Figure 11 (see page 70) shows the relative patterns of mean-factor scores for students falling in each group, while Figure 12 (see page 70) summarises where significant differences between these were found.

As can be seen from both figures, learners at the C level of proficiency have the highest scores on all scales except parental encouragement. These differences are a significant increase on the B2 group for four scales: English self-concepts, ideal L2 self, language learning experience and motivated behaviour. Thus, what particularly distinguishes these learners, compared to those in other groups, is that they are most confident about their ability to learn English, have the most elaborate and plausible vision of themselves as successful users of the English language in the future, show the most positive language learning attitudes, and invest most effort in language learning.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W 7 4 1 6 B E N G L I S H ? I M P A C T B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H S 8 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Figure 11: Motivational profiles for learners at different levels of proficiency

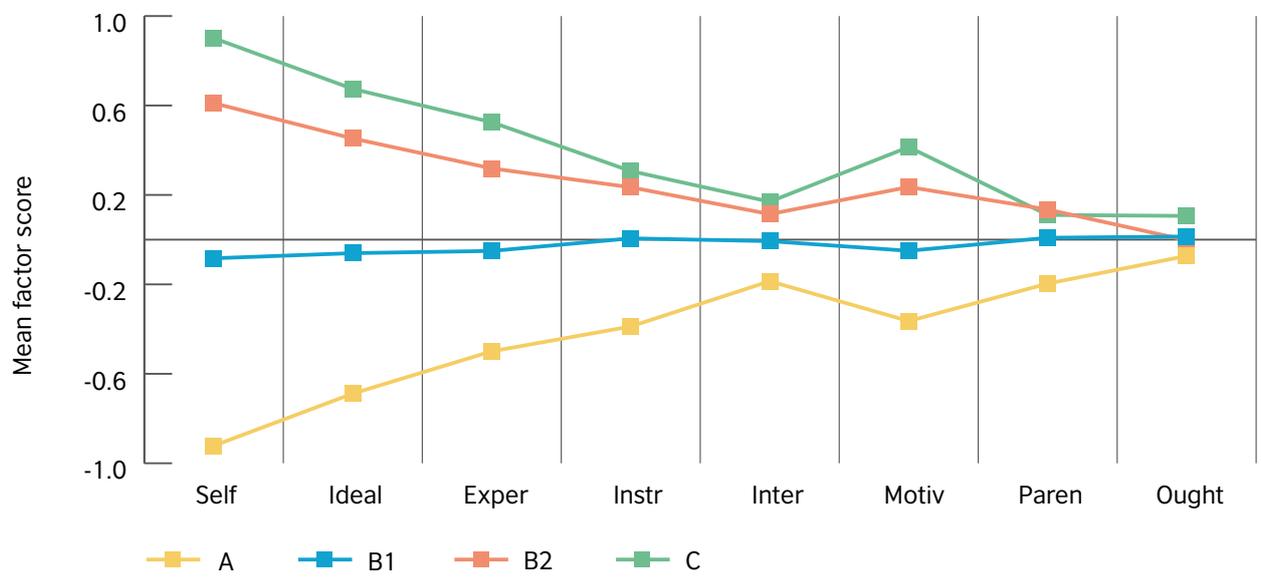


Figure 12: Visual representation of significant differences on motivational scales for learners at different levels of proficiency

CEFR level	Motivational scale								
	English self-concept	Ideal L2 self	Language learning experience	Instrumentality	International orientation	Motivated behaviour	Parental encouragement	Ought-to L2 self	
C	Dark Green	Dark Green	Dark Green	Light Green	Light Green	Dark Green	Dark Green	Light Green	Light Green
B2	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green
B1	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green
A	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green	Very Light Green

Key: darker colour = higher factor score; different shades = statistically significant difference

The overall pattern of motivational traits for learners at B2 level of proficiency seen in Figure 12 reflects that of the learners at C level. In fact, the values do not differ significantly from learners at C level on scales of instrumentality, international orientation and parental encouragement. This means that

they see English as equally useful for future communication and in their professional lives and report a similar amount of parental encouragement as learners at C level.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 7 5 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Meanwhile, learners at B1 level of proficiency score significantly lower on all scales but ought-to L2 self than learners at B2 level, and higher on all scales but ought-to L2 self than learners at A level. The pattern of motivations among these B1-level learners can be characterised as low to moderate in all aspects, lacking the spike seen for the higher-level learners on the self-concept and ideal L2-self scales.

Students at the A level of proficiency have the lowest scores on all variables. Further analysis confirms that these differences are significant on all but one scale (the ought-to L2 self).

One point to note is that the scale of the ought-to L2 self has not been found to play a decisive role in determining students' profiles. There was only one significant difference between students at C and A level, with the latter group more resistant to the expectations imposed on them by the external environment than the C-level group. This is in line with previous findings reported here, which show that the relationship between the ought-to L2 self and proficiency is weak or non-existent.

On the whole, however, the motivational scores on all but this one scale follow the patterns of proficiency. Higher motivation is associated with higher proficiency: the more proficient the students, the more confident and goal-oriented they are, the more positive attitudes they have, the more effort they invest in language learning and they report greater levels of parental encouragement.

Correlation analysis

The results presented in Table 34 show the strength of correlations between the overall scale score⁷ and the factor scores for motivational scales. The correlations are listed in descending strength.

Table 34: Correlations between motivational variables and overall total score

Motivational variables	Total score
English self-concept	.652**
Ideal L2 self	.567**
Language learning experience	.421**
Instrumentality	.412**
International orientation	.375**
Motivated behaviour	.303**
Parental encouragement	.216**
Ought-to L2 self	.050*

* $p < .05$ ** $p < .01$

7. The total score achieved by the participants in the four skill areas, out of a possible total of 200 points (50 per component). In this dataset, total scale scores ranged between 22 and 196.

As can be seen, all motivational variables are positively related to proficiency, although the correlation between ought-to L2 self and the overall scale is very weak. There are large differences in correlation strength. This can be clearly observed when comparing scatter plots showing the bivariate relationship between factor scores for the scale most closely related to proficiency (English self-concept, shown in Figure 13) and least closely related to proficiency (the ought-to L2 self, shown in Figure 14). In the first case, it is clearly visible that the scores on English self-concept grow in tandem with the overall proficiency scores, whereas in the case of the ought-to L2 self the points are much more widely dispersed.

In terms of assessing the strength of correlations, there are two strong correlations with proficiency (English self-concept, ideal L2 self), four medium correlations (language learning experience, instrumentality, international orientation and motivated behaviour) and one small correlation (parental encouragement) (interpretation according to Cohen, 1988). The order of strength of correlation mirrors the results from the covariances between latent variables estimated in the confirmatory factor analysis described above, with variables displaying the strongest relationships with proficiency representing the more internalised motives for studying English, and those with the weaker relationships more externalised.

Figure 13: Correlation between overall proficiency and English self-concept

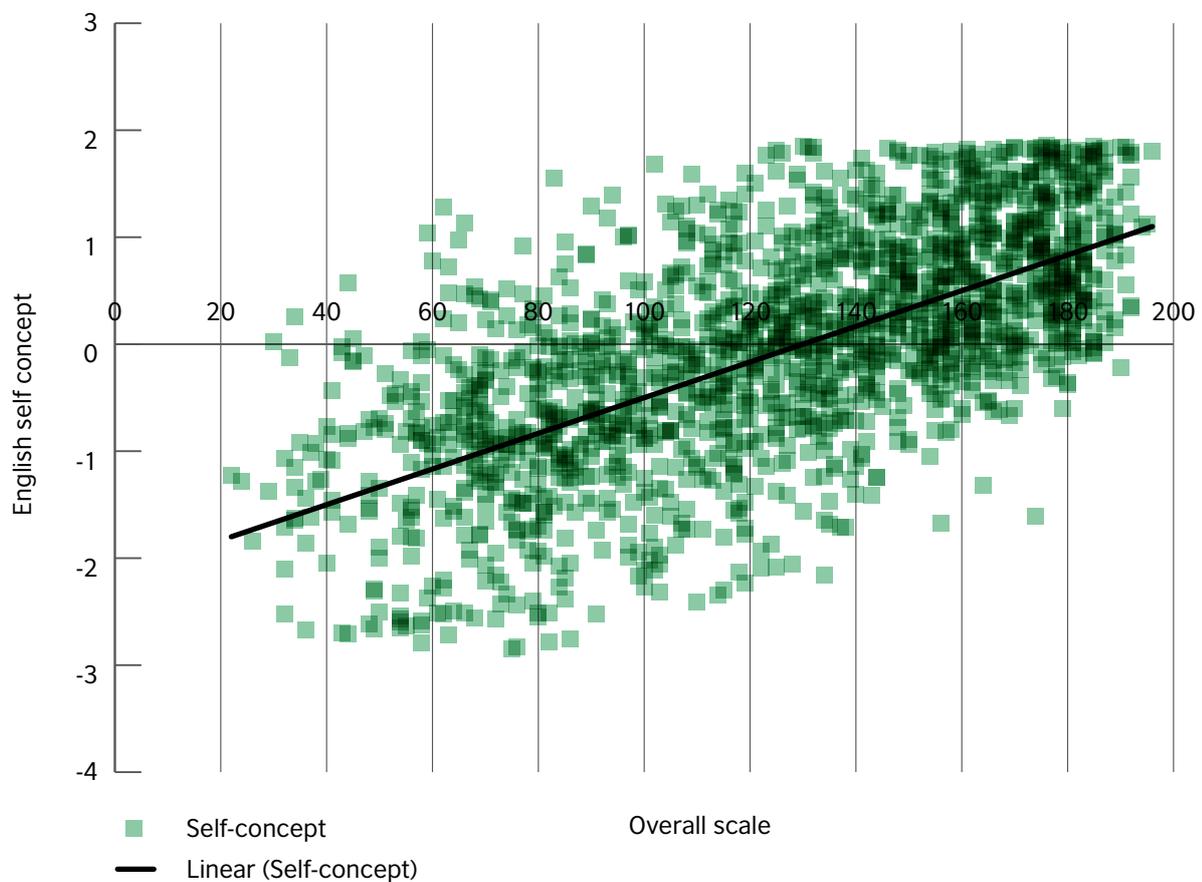


Figure 14: Correlation between overall proficiency and the ought-to L2 self

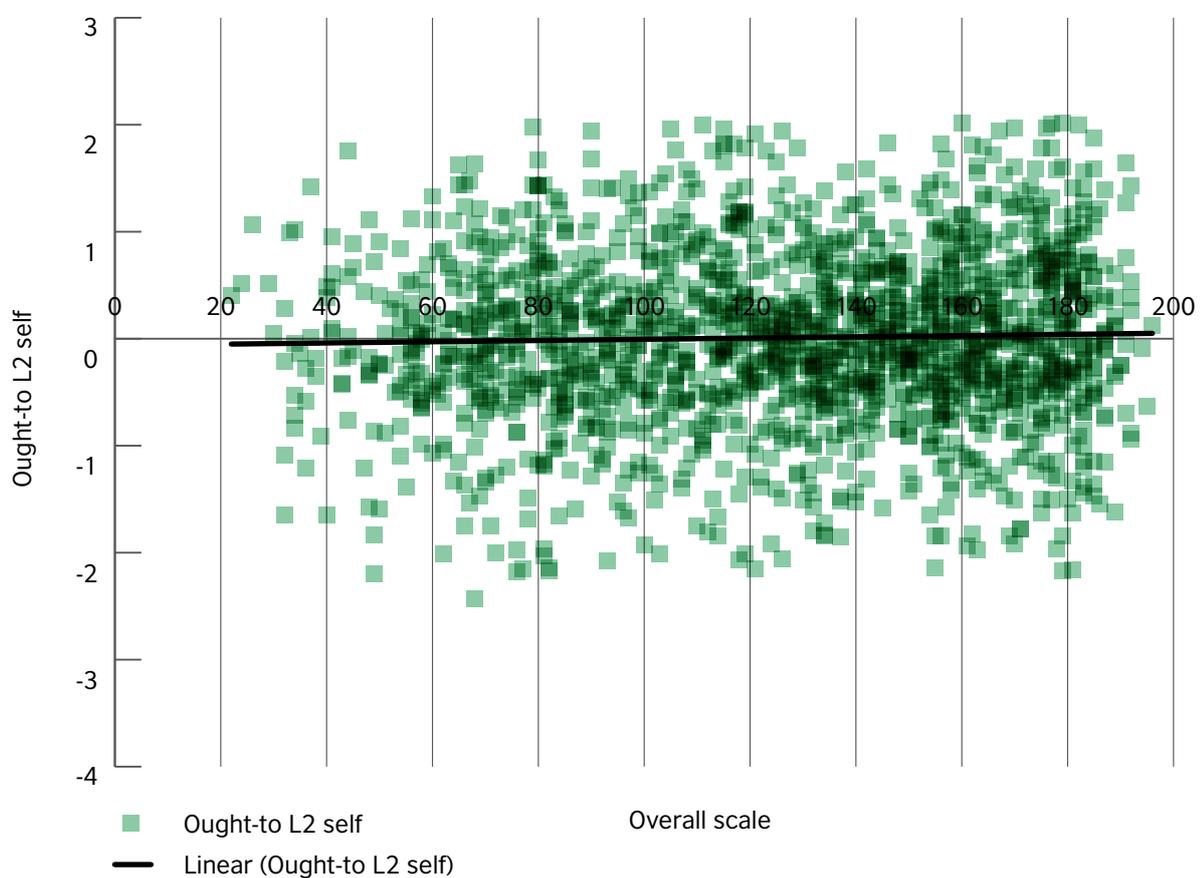


Table 35 (see page 78) shows correlations between motivational variables and components of the proficiency test. Compared to the overall proficiency scores, correlations with scores on individual components are slightly weaker. This is because the measure of overall proficiency is a composite of skills; hence, it is a more comprehensive scale.

Correlations with speaking, grammar and vocabulary tend to be slightly higher than with other parts of the test, whereas correlations with reading tend to be the lowest. These differences are, however, small, and the order of strength of correlation remains the same for all parts of the test. In the case of the ought-to L2 self, correlations with reading and listening are not statistically significant. This suggests that the pressure to learn English from the external environment bears no relationship with scores on the reading and listening parts of the test.

Table 35: Correlations between motivational variables and scale scores for individual Aptis components

Motivational scales	Listening	Reading	Speaking	Writing	Grammar and vocabulary
English self-concept	.569**	.598**	.570**	.596**	.620**
Ideal L2 self	.514**	.504**	.509**	.505**	.533**
Language learning experience	.365**	.374**	.388**	.379**	.388**
Instrumentality	.363**	.338**	.397**	.376**	.352**
International orientation	.344**	.309**	.359**	.335**	.329**
Motivated behaviour	.257**	.256**	.288**	.283**	.281**
Parental encouragement	.187**	.162**	.230**	.194**	.172**
Ought-to L2 self	.025	.017	.075**	.060*	.024

* $p < .05$ ** $p < .01$

Below is the discussion of findings scale by scale.

- English self-concept – this is most closely related with the measures of proficiency. This means that students from Madrid make generally correct evaluations of their own ability to learn and speak English. This finding is unsurprising, as previous studies (Hsieh and Kang, 2010; Hsieh and Schallert, 2008; Mills et al., 2007) reported a link between self-concept and achievement. As well as reflecting performance, it is also understood that ensuring students' English self-concept is positive can be a first step towards higher proficiency in a circular cause-and-effect relationship (Bandura, 1997).

- Ideal L2 self – this is also strongly correlated with English proficiency. This means that students who have a robust vision of themselves as future successful users of English tend to have higher proficiency levels than their peers without such a vision. In order to build such a vision, it is vital that learners have time to reflect on what they will be using English for in the future. This finding is in contrast to the finding reported by Moskovsky et al. (2016). However, as numerous studies reported a link between effort investment and the ideal L2 self (Iwaniec, 2014; Kormos et al., 2011; Taguchi et al., 2009), it is unsurprising.

- Language learning experience – positive language learning experience, measured by attitudes to language learning, is moderately conducive to proficiency in English. A possible explanation for this might be that interesting and rewarding learning experiences can help sustain effort investment in the short term (Ford, 1992) and create a positive outlook for future activities. Hence, it is vital that English classes are a source of such positive experiences.
- Instrumentality and international orientation – having clear language learning goals, such as using English in a future career (instrumentality) or being able to communicate with other English speakers around the world (international orientation), is moderately correlated with students' proficiency. This is because having a language learning goal helps direct effort (Ford, 1992). Learners might be more likely to adopt language learning goals if they are presented with positive role models. This might, for example, take the form of short meetings with local people discussing how English is useful in their jobs (for example, small business owners, builders, and those who work or worked abroad). To raise the profile of English as a language of communication, opportunities for communication might be created (for example, exchanges or trips abroad). Equally, finding a twin school from abroad to communicate with via the internet might be very beneficial.
- Motivated learning behaviour – this is also moderately related to proficiency. This means that there is a link between the reports of effort invested in language learning and student proficiency.

To make the link stronger, it is crucial that learners have opportunities to familiarise themselves and reflect on their learning styles. It is also worth mentioning that there might have been substantial fluctuations in effort investment during the whole period of study. Hence, the measure of effort at one point in time might not be sufficient to fully explain the correlation between effort investment and motivation.

- Parental encouragement – the correlation between parental encouragement and proficiency is positive but low. This is unsurprising, considering that students from the current study are 15-year-olds, which is a stage when teenagers value newly gained independence. Whereas previous studies did not look specifically at the relationship between parental encouragement and students' proficiency, the existing research shows that parental encouragement has a limited influence on students' motivation (Iwaniec and Ullakonoja, 2016; Iwaniec, 2015; Kyriacou and Zhu, 2008; Lamb, 2012), which is likely to affect proficiency.
- Ought-to L2 self – the link between the ought-to L2 self and proficiency is very weak. This means that the pressure from the external environment is not a force motivating students to learn English. This finding is unsurprising, as previous studies have pointed to the limited role of the ought-to L2 self in European learners (Csizér and Kormos, 2008a, 2008b, 2009).

The following analysis examines where differences lie in terms of these correlations for the different student groupings, split by gender, schooling and SES as per the analysis reported above.

Correlations by female and male populations

The results of correlational analysis of the cohort split along gender lines (Table 36) show there are some significant differences in the strength of relationships between motivational variables and the overall results on the

proficiency test on two motivational scales: language learning experience and international orientation.

These differences are mirrored when correlations with individual components of the test are explored.

Table 36: Correlations between overall proficiency and motivational variables, by gender

Motivational scale	Female	Male	<i>p</i> -value [†]
English self-concept	.668	.631	.177
Ideal L2 self	.563	.568	.880
Language learning experience	.463	.379	.032*
Instrumentality	.402	.417	.704
International orientation	.414	.334	.050*
Motivated behaviour	.294	.302	.818
Parental encouragement	.179	.257	.085
Ought-to L2 self	.084	.078	.897

[†]Based on Fisher's z-score **p* < .05

In detail:

- Language learning experience – scores on language learning experiences are more closely related to proficiency for females than for males. This means that females are more likely to translate positive language learning attitudes into an increase in proficiency than males. Hence, it could be concluded that males experience some obstacles along the way (Ajzen, 2005). The reasons might be multiple. For example, inflated self-concept (see above discussion on levels of motivation across gender groups), a poor match of language classes with boys’ preferred learning styles (Carr and Pauwels, 2005a), or in the use of language learning strategies (Oxford, 1994).
- International orientation – females’ scores on international orientation are more closely related to proficiency than males’ scores. This suggests that girls are more strongly oriented towards communication – a result that has

been previously reported by Chávez (2000), who noticed that females have a stronger preference for communication and valuing language proficiency than males. Similarly, Henry and Cliffordson (2013) concluded that females are more communication-oriented than males.

Correlations by socio-economic status factor scores

The motivational data for the higher SES group⁸ generally appears to be more closely related to proficiency than for the lower SES group in the case of all but one variable, namely English self-concept. The results of the correlational analysis of motivational variables and overall proficiency are presented in Table 36. The differences in the strength of correlation between motivation and proficiency are significant in the case of three variables: international orientation, motivated behaviour and the ought-to L2 self. In more detail:

Table 37: Correlations between overall proficiency and motivational variables by SES

Motivational scale	Lower SES	Higher SES	p-value [†]
English self-concept	.629	.630	.976
Ideal L2 self	.523	.574	.126
Language learning experience	.394	.452	.136
Instrumentality	.352	.416	.114
International orientation	.327	.417	.028*
Motivated behaviour	.239	.362	.004**
Parental encouragement	.148	.214	.152
Ought-to L2 self	.002	.099	.040**

[†]Based on Fisher’s z-score *p< .05 **p<.01

- International orientation – learners with a higher SES who want to use English to communicate with others appear to have higher proficiency than students from lower SES backgrounds with the same goal. One explanation for such a difference could be lower levels of English self-concept among the latter group. Ford (1992) explains that if goals are not considered attainable, they lose their motivational force. As students with lower SES backgrounds consider themselves less capable in learning and using English than their more advantaged peers (see above discussion on findings from the multi-group analysis), their adopted goals might lose part of their motivational power. Moreover, students from the higher SES group might have more opportunity to use English for communication when travelling abroad than their peers from the lower SES group, making this language learning goal more desirable.
- Motivated learning behaviour – effort invested in language learning by members of the higher SES group is more closely related to their proficiency than for students from the lower SES group. It needs to be acknowledged that the results suggest this is the biggest difference in correlation with proficiency between the two groups. This suggests there might be a qualitative difference to the effort between the two groups with the higher SES group employing more efficient language learning strategies than their peers from the

lower SES group. This might be counteracted by focusing instruction in school more on explicitly teaching how to learn effectively and providing individual help to students to better understand how they learn.

- Ought-to L2 self – while there is a small but significant correlation between higher SES students and their proficiency, the correlation for lower SES students is not significant. This means that the pressure to learn English felt by students in the lower SES group is not in any way linked to their ultimate proficiency. In contrast, for higher SES students, the role of this pressure is positive, albeit small.

Correlations by student populations from bilingual and non-bilingual schools

Bilingual schools differ from non-bilingual schools in the type and amount of exposure to the English language that students receive. Whereas in non-bilingual schools students' English instruction is generally limited to English language lessons, students in bilingual schools are exposed to English in other subjects. Thus, students at bilingual schools have more exposure to the English language, and as well as being a subject of study it is also experienced as a medium of instruction. As the students engage with the English language in a different way, it is expected that this will affect the strength of the relationship between motivation and proficiency.

8. See section (results/multi-group/SES) above for description of how the SES groups were derived.

Table 38: Correlations between overall proficiency and motivational variables, by school type

Motivational scale	Bilingual schools	Non-bilingual schools	p-value [†]
English self-concept	.623	.655	.298
Ideal L2 self	.587	.544	.226
Language learning experience	.394	.432	.380
Instrumentality	.346	.446	.023**
International orientation	.389	.376	.772
Motivated behaviour	.304	.299	.912
Parental encouragement	.166	.280	.056
Ought-to L2 self	.030	.138	.037**

[†]Based on Fisher’s z-score **p<.01

Overall, it can be noticed that the correlations between motivation and proficiency tend to be stronger for learners from non-bilingual rather than bilingual schools. The key here might be the amount of exposure to English these two groups receive. As students from bilingual schools receive much more exposure, motivation plays a less important role in learning. This is because a larger proportion of learning might happen subconsciously (for example, the learner subconsciously revises vocabulary and practises listening skills when the teacher teaches biology in English) as compared to learning in non-bilingual schools, which tends to be limited to English language classes.

As can be seen in Table 39, there are two significant differences in the strength of correlations. These are between proficiency and instrumentality and ought-to L2 self. In these two cases, the link with proficiency is higher for students from non-bilingual schools, rather than for bilingual schools. In more detail:

- instrumentality – students from non-bilingual schools who have professional plans involving English are more likely to have higher scores on the proficiency tests than their peers from bilingual schools

- ought-to L2 self – whereas the correlation between the ought-to L2 self and proficiency is not significant, there is a positive but very weak relationship between societal expectations and proficiency.

These findings point to the conclusion that the motivation of learners from non-bilingual schools is more dependent on external factors, such as social expectations, or even future career opportunities, and these factors explain more variance in the achievement of this group as compared to learners from bilingual schools.

A closer look at correlations for non-bilingual schools

The non-bilingual schools can be further split into state-funded schools (n=574) and charter schools (n=672). As can be seen in Table 38 the strength of correlation between motivational variables and proficiency is generally higher for students from charter schools than those from non-bilingual schools, with the exception of motivated behaviour. However, the differences are not as great as those between bilingual and all non-bilingual schools, and none of them reaches significance.

Table 39: Correlations between overall proficiency and motivational variables for non-bilingual schools by public/charter status

Motivational scale	State-funded non-bilingual schools	Charter non-bilingual schools	p-value [†]
English self-concept	.655**	.673**	.569
Ideal L2 self	.508**	.554**	.259
Language learning experience	.439**	.443**	.897
Instrumentality	.400**	.441**	.818
International orientation	.330**	.398**	.168
Motivated behaviour	.331**	.289**	.412
Parental encouragement	.212**	.256**	.412
Ought-to L2 self	.099**	.092*	.903

[†]Based on Fisher's z-score * $p < .05$ ** $p < .01$

The findings from this chapter show that motivation is clearly related to proficiency, with the English L2 self-concept – understood as confidence in language learning – found to be the variable most closely related to achievement. Yet learners from non-bilingual schools and from lower socio-economic backgrounds reported lower levels of self-assurance than their peers. Similar findings from non-bilingual schools were reported in the case of other influential variables, such as the ideal L2 self, instrumentality, international orientation, language learning experience and motivated behaviour. However, parental expectations were found to be of limited influence at this stage of education, with external motivators also playing a negligible role. This means that, overall, 15-year-olds from the Madrid region are largely acting on internalised English language learning motivations. A closer look at group-level analyses reveals some interesting distinctions.

Gender-wise, males reported overall lower levels of motivation than females. They also appear to rely more than females on external stimuli, such as parental and social expectations, to motivate them. This could be interpreted to be linked to general perceptions of language learning as a female domain, with language proficiency perceived as more directly relevant for future careers of female than male learners. Despite higher overall proficiency scores across the board, it is interesting to note that female learners did not report themselves to be better at learning English than their male counterparts. This is consistent with previous research findings, which imply that females tend to be more conservative in their judgement of ability.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 8 5 3 Q J H
G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Splitting students by a measure of socio-economic status derived from survey responses was found to provide the most consistent division in the dataset in terms of motivational levels. Students from the lower socio-economic group reported lower levels of motivation in all areas. Moreover, the influence of some aspects of motivation on achievement was weaker for this group of students. In particular, the outward-looking international orientation variable was found to be less clearly related with higher achievement, as was the role of societal pressure. In addition, motivated learning behaviour was less likely to be translated into better test performance. The reasons behind the distinctions in these relationships are multifarious and complex. However, a key finding of the analysis was to show that regardless of socio-economic status, attendance at a bilingual school brings gains in both proficiency and positive underlying motivations.

Learning English in schools participating in the Bilingual programme appears not only to lead to higher achievement but also to the development of healthy language learning motivation. The findings indicate that these students have internalised the value of learning English to a greater extent than students at Non-Bilingual schools, with greater levels of confidence and more robust visions of themselves as successful English users in the future. Their experience of communicating in English on a day-to-day basis gives them immediate evidence for the role of English for communicative purpose, with evidence suggesting a lower need amongst these students to look to external impetus to perform competently in English. Certainly, significantly lower correlations with test performance for the instrumentality and obligation variables suggests that these students are making proficiency gains without feeling the same pressures as their peers attending Non-Bilingual schools.

9. IN CONCLUSION

The research aims outlined at the beginning of this report established the intension to evaluate the English language capability of students studying at public schools within the Madrid region of Spain; to compare the outcomes in schools participating in the region's Bilingual programme with those not yet participating in the programme, and to understand the relationship between English language learning motivation and increased proficiency. An evaluation of capability, as described in the introductory chapter, involves the assessment of both the level of achievement reached by a defined population; and the opportunities provided to them to achieve greater proficiency via teaching and learning practice derived from a policy or national guideline. The aim of this research was not to create a single score to show the success, or otherwise, of the policy for English language teaching and learning in the Madrid region of Spain. The detailed description of the research processes and outcomes presented in this report are intended to provide policymakers, teachers and researchers with a full analysis of the complex and intertwined elements that combine to influence students' learning outcomes.

Three central factors came together to produce this groundbreaking research. The research design; combining ACER's sampling expertise and the British Council's knowledge and experience in English language assessment enabled a collaboration leading to world-class research outcomes. Complex field operations within schools often provide significant challenges to successful completion of large-scale data collection, therefore the consistent use of offline-enabled tablet computers to deliver all language assessments in every class room has contributed to the success of English Impact Madrid. Lastly, the positive collaboration with teachers, schools and students participating in the research process has been invaluable. Without their positive and proactive co-operation this evaluation would not have produced the strong and reliable evidence upon which further discussions and policy decisions may be based.



REFERENCES

- Ajzen, I (2005) *Attitudes, personality and behaviour*. New York: Open University Press.
- Bachman, LF (1990) *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bandura, A (1997) *Self-efficacy: The exercise of control*. New York: WH Freeman & Company.
- Bong, M and Skaalvik, EM (2003) Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review* 15/1: 1–40.
- Brown, TA (2015) *Confirmatory Factor Analysis for Applied Research* (2nd Edition). New York, NY: Guildford Press
- Byrne, BM (2012) *Structural equation modeling with Mplus: Basic concepts, applications and programming*. Routledge.
- Caro, DH and Cortés, D (2012) Measuring family socioeconomic status: An illustration using data from PIRLS 2006. Available online at: www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_05_Chapter_1.pdf \t “_blank”
- Carr, J and Pauwels, A (2005a) *Boys and foreign language learning*. Basingstoke: Palgrave Macmillan.
- Carr, J and Pauwels, A (2005b) *Boys and foreign language learning: Real boys don't do languages*. Basingstoke: Palgrave Macmillan.
- Chávez, M (2000) *Gender in the language classroom*. London: McGraw Hill.
- Clark, A (1998) *Gender on the agenda: Factors motivating boys and girls in MFLs*. London: Centre for Information on Language Teaching and Research.
- Cohen, J (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale Lawrence Erlbaum Associates.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Csizér, K and Kormos, J (2008a) Modelling the role of inter-cultural contact in the motivation of learning English as a foreign language. *Applied Linguistics* 30/2: 166–185. Available online at: <http://doi.org/10.1093/applin/amn025>
- Csizér, K and Kormos, J (2008b) The relationship of intercultural contact and language learning motivation among Hungarian students of English and German. *Journal of Multilingual and Multicultural Development* 29/1: 30–48. Available online at: <http://doi.org/10.2167/jmmd557.0>
- Csizér, K and 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 Dörnyei, Z and Ushioda, E (eds) *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters, 98–119.

Davis-Kean, PE (2005) The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology* 19/2: 294–304. Available online at: <http://doi.org/10.1037/0893-3200.19.2.294>

Dörnyei, Z and Ushioda, E (eds) (2009) *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters.

Eccles, JS (1993) 'School and family effects on the ontogeny of children's interests, self-perceptions, and activity choice', in Jacobs, J (ed) *Nebraska Symposium on Motivation: Vol. 40. Developmental perspectives on motivation*. Lincoln University, 145–208.

Eccles, JS (1994) Understanding women's educational and occupational choices. Applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly* 18: 585–609. Available online at: <http://doi.org/10.1111/j.1471-6402.1994.tb01049.x>

Edele, A, Seuring, J, Kristen, C and Stanat, P (2015) Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research* 52: 99–123. Available online at: <http://doi.org/10.1016/j.ssresearch.2014.12.017>

Erling, EJ and Seargeant, P (2013) *English and development: Policy, pedagogy and globalization*. Bristol: Multilingual Matters.

Fernandez Fontecha, A (2010) 'Gender and motivation in EFL vocabulary production', in Jimenez Catalan, RM (ed) *Gender perspectives on vocabulary in foreign and second languages*. Basingstoke: Palgrave Macmillan, 93–116

Ford, M (1992) *Motivating humans: Goals, emotions, and personal agency beliefs*. London: Sage Publications.

Gardner, RC and Lambert, WE (1959) Motivational variables in second language acquisition. *Canadian Journal of Behavioural Psychology* 13: 266–72.

Gardner, RC and Lambert, WE (1972) 'Motivational variables in second-language acquisition', In Gardner, RC and Lambert, WE (eds) *Attitudes and motivation in second-language learning*. Rowley: Newsbury House.

Graddol, D (2006) *English next: Why Global English may mean the end of 'English as a foreign language.'* London: British Council.

Henry, A and Cliffordson, C (2013) Motivation, gender, and possible selves. *Language Learning* 63/2: 271–295. Available online at: <http://doi.org/10.1111/lang.12009>

Hsieh, PP-H and Kang, HS (2010) Attribution and self-efficacy and their interrelationship in the Korean EFL context. *Language Learning* 60/3: 606–627. Available online at: <http://doi.org/10.1111/j.1467-9922.2010.00570.x>

Hsieh, PP-H and Schallert, DL (2008) Implications from self-efficacy and attribution theories for an understanding of undergraduates' motivation in a foreign language course. *Contemporary Educational Psychology* 33/4: 513–532. Available online at: <http://doi.org/10.1016/j.cedpsych.2008.01.003>

IBM Corp. (2013) IBM SPSS Statistics for Windows. Armonk: IBM Corp.

Iwaniec, J (2014) Motivation of pupils from southern Poland to learn English. *System* 45/1: 67–78. Available online at: <http://doi.org/10.1016/j.system.2014.05.003>

Iwaniec, J (2015) *Motivation to learn English of Polish gymnasium pupils*. Lancaster University. Available online at: <http://opus.bath.ac.uk/49838/>

Iwaniec, J and Ullakonoja, R (2016) Polish and Finnish teenagers' motivation to learn English: The role of context. *European Journal of Applied Linguistics* 4/2: 277–300. Available online at: <http://doi.org/10.1515/eujal-2015-0034>

Jiménez Catalán, RM (2010) 'Gender tendencies in lexical acquisition and use', in Jiménez Catalán, RM (ed) *Gender perspectives on vocabulary in foreign and second languages*. Basingstoke: Palgrave Macmillan, 117–138

Kim, T-Y and Kim, Y-K (2014) 'EFL students' L2 motivational self system and self-regulation: Focusing on elementary and junior high school students in Korea' in Csizér, K and Magid, M (eds) *The impact of self-concept on language learning*. Bristol: Multilingual Matters, 87–107

Klebanov, PK, Brooks-Gunn, J and Duncan, GJ (1994) Does neighborhood and family poverty affect mothers' parenting, mental health, and social support. *Journal of Marriage and Family* 56/2: 441–455. Available online at: www.jstor.org/stable/353111

Kormos, J and Kiddle, T (2013) The role of socio-economic factors in motivation to learn English as a foreign language: The case of Chile. *System* 41/2: 399–412. Available online at: <http://doi.org/10.1016/j.system.2013.03.006>

Kormos, J, Kiddle, T and Csizér, K (2011) Systems of goals, attitudes, and self-related beliefs in second-language learning motivation. *Applied Linguistics* 32/5: 495–516. Available online at: <http://doi.org/10.1093/applin/amr019>

Kyriacou, C and Zhu, D (2008) Shanghai pupils' motivation towards learning English and the perceived influence of important others. *Educational Studies* 34/2: 97–104. Available online at: <http://doi.org/10.1080/03055690701811099>

Lamb, M (2012) A self system perspective on young adolescents' motivation to learn English in urban and rural settings. *Language Learning* 62/4: 997–1,023. Available online at: <http://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* 34/1: 1–16. Available online at: <http://doi.org/10.1080/01434632.2012.697467>

Marsh, HW (1990) The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology* 82/4: 623–636. Available online at: <http://doi.org/10.1037//0022-0663.82.4.623>

Marsh, HW and Martin, A J (2011) Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology* 81: 59–77. Available online at: <http://doi.org/10.1348/000709910X503501>

Mills, N, Pajares, F and Herron, C (2007) Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning* 57/3: 417–442. Available online at: <http://doi.org/10.1111/j.1467-9922.2007.00421.x>

Moskovsky, C, Racheva, S, Assulaimani, T and Harkins, J (2016) The L2 Motivational Self System and L2 Achievement: A Study of Saudi EFL Learners. *The Modern Language Journal* 16: 1–14. Available online at: <http://doi.org/10.1111/modl.12340>

Muthén, LK and Muthén, BO (2011) *Mplus user's guide* (6th ed.). Los Angeles: Muthén & Muthén.

Norton, B and Pavlenko, A (2004). 'Gender and English language learners: Challenges and possibilities', in Norton, B and Pavlenko, A (eds), *Gender and English language learners* (pp. 1–12). Virginia: TESOL.

O'Sullivan, B (2011) 'Language testing', in Simpson, J (ed), *Routledge handbook of applied linguistics*. Oxford: Routledge.

O'Sullivan, B (2015) *Aptis test development approach*. Aptis Technical Report, TR/2015/001. London: British Council.

O'Sullivan, B and Weir, CJ (2011) 'Language testing and validation', in O'Sullivan, B (ed) *Language testing: theory & practice*. Oxford: Palgrave, 13–22.

O'Sullivan, B and Dunlea, J (2015). *Aptis General Technical Manual V 1.0*. Aptis Technical Report, TR/2015/004. London: British Council.

OECD (2013) *PISA 2012 results: Excellence through equity: Giving every student the chance to succeed* (Volume II). OECD Publishing. Available online at: <http://doi.org/http://dx.doi.org/10.1787/9789264201132-en>

Ortega, L (2009) *Understanding second language acquisition*. London: Hodder Education.

Oxford, R (1993) Instructional implications of gender differences in second/foreign language (L2) learning styles and strategies. *Applied Language Learning* 5: 65–94.

Oxford, R (1994) 'La difference continue... Gender differences in second/foreign language learning styles and strategies', in Sunderland, J (ed), *Exploring gender: Questions and implications for English language education*. Prentice Hall, 140–147.

Pfenninger, SE and Singleton, D (2016) Affect trumps age: A person-in-context relational view of age and motivation in SLA. *Second Language Research* 32/3: 311–345. Available online at: <http://doi.org/10.1177/0267658315624476>

Ryan, RM and Deci, EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55/1: 68–78. Available online at: <http://doi.org/10.1037/110003-066X.55.1.68>

Saito, M (2003) Amartya Sen's Capability Approach to education: a critical exploration. *Journal of Philosophy of Education* 37/1: 17–33.

Satorra, A and Bentler, PM (2010) Ensuring positiveness of the scaled difference chi-square test statistics. *Psychometrika* 75/12: 243–248. Available online at: <http://doi.org/doi:10.1007/s11336-009-9135-y>

Shavelson, RJ, Hubner, JJ and Stanton, GC (1976) Self-concept: Validation of construct interpretations. *Review of Educational Research* 46/3: 407–441.

Shohamy, E (1998) Critical language testing and beyond. *Studies in Educational Evaluation* 24/4: 331–345.

Skaalvik, S and Skaalvik, EM (2004) Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles* 50/3–4: 241–252. Available online at: <http://doi.org/10.1023/B:SERS.0000015555.40976.e6>

Taguchi, T, Magid, M and Papi, M (2009) 'The L2 motivational self-system among Japanese, Chinese and Iranian learners of English: A comparative study', in Dörnyei, Z and Ushioda, E (eds), *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters, 66–98.

Taylor, F and Marsden, EJ (2014) Choosing to study foreign languages in England: An experimental intervention. *The Modern Language Journal* 98/4: 902–920. Available online at: <http://doi.org/10.1111/modl.12146>

Ushioda, E (2009) 'A person-in-context relational view of emergent motivation, self and identity', in Dörnyei, Z and Ushioda, E (eds), *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters, 215–228.

Weir, CJ (2005) *Language Testing and Validation: an evidence-based approach*.

Williams, M and Burden, RL (1997) *Psychology for language teachers: A social constructivist approach*. Cambridge, UK: Cambridge University Press.

Yashima, T (2009) 'International posture and the ideal L2 self in the Japanese EFL context', in Dörnyei, Z and Ushioda, E (eds), *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters, 144–164.

A :: FNOAG//XCKPW?9BE//17B=6>?9BEB17B=ZV28TZQU-<2535%ANOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17? ?//GWBSpain, Madrid9>?//GWB::17B=6PW8FW?4>JHY4H933QJHG%2AXGML1981^9#DRM<L%A9M5<ZV28TIQU<2535%AXG%L1UM8FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?



APPENDIX A: OVERVIEW OF THE APTIS FOR TEENS TEST STRUCTURE

Structure of the teens core component

Part	Skill focus	Items/ part	Level	Tasks/ level	Items/ task	Task focus	Task description	Response format
Test: Core (50 items)								
1	Grammar	25	A1	5	1	Syntax and word usage	Sentence completion: select the best word to complete a sentence based on syntactic appropriacy.	Three-option multiple choice.
			A2	5–7	1			
			B1	5–7	1			
			B2	5–7	1			
2	Vocabulary	25	A1	1	5	Synonym (vocabulary breadth)	Word matching: match two words which have the same or very similar meanings.	Five target words. Select the best match for each from a bank of ten options.
			A2	1	5	Meaning in context (vocabulary breadth)	Sentence completion: select the best word to fill a gap in a short sentence. Understanding meaning from context.	Five sentences, each with a one-word gap. Select the best word to complete each from a bank of ten options.
			B1	1	5	Meaning in context (vocabulary breadth)	Sentence completion: select the best word to fill a gap in a short sentence. Understanding meaning from context.	Five sentences, each with a one-word gap. Select the best word to complete each from a bank of ten options.
				1	5	Definition (vocabulary breadth)	Matching words to definitions.	Five definitions. Select the word defined from a bank of ten options.
B2	1	5	Collocation (vocabulary depth)	Word matching: match the word that is most commonly used with a word targeted from the appropriate vocabulary level.	Five target words. Select the best match for each from a bank of ten options.			

Structure of the teens reading component

Part	Skill focus	Items/ part	Level	Tasks/ level	Items/ task	Task focus	Task description	Response format
Test: Reading (25 items)								
1	Sentence-level meaning	5	A1	1	5	Sentence-level meaning (careful, local reading)	Gap fill. A short text with five gaps. Filling each gap only requires comprehension of the sentence containing the gap. Text-level comprehension is not required.	Three-option multiple choice for each gap.
2	Inter-sentence cohesion	6	A2	1	6	Inter-sentence cohesion (careful global reading)	Re-order jumbled sentences to form a cohesive text.	Re-order six jumbled sentences. All sentences must be used to complete the story.
3	Text-level comprehension of short texts	7	B1	1	7	Text-level comprehension of short texts (careful global reading)	Candidates match four short paragraphs giving information about four people's opinions on different topics and identify which of the four people could say certain statements.	Seven gaps in a short text. Select the best word to fill each gap from a bank of nine options.
4	Text-level comprehension of long text	7	B2	1	7	Text-level comprehension of longer text (global reading, both careful and expeditious)	Matching the most appropriate headings to paragraphs. Requires integration of micro- and macro-propositions within and across paragraphs, and comprehension of the discourse structure of more complex and abstract texts.	Seven paragraphs forming a long text. Select the most appropriate heading for each paragraph from a bank of eight options.

Structure of the teens listening component

Skill focus	Items/ part	Level	Tasks/ level	Items/ task	Format	Task description	Response format
Test: Listening (25 items)							
Lexical recognition	10	A1	10	1	Monologues	Q&A about listening text. Listen to short monologues (recorded messages) to identify specific pieces of information (numbers, names, places, times, etc.).	Three-option multiple choice. Only the target is mentioned in the text.
Identifying specific, factual information	5	A2	5	1	Monologues and dialogues	Q&A about listening text. Listen to short monologues and conversations to identify specific pieces of information (numbers, names, places, times, etc.).	Three-option multiple choice. Lexical overlap between distractors and words in the input text.
Identifying specific, factual information	5	B1	5	1	Monologues and dialogues	Q&A about listening text. Listen to short monologues and conversations to identify propositions. The information targeted is concrete and of a factual/literal nature. Requires integration of information over more than one part of the input text.	Three-option multiple choice. Distractors should have some overlap with information and ideas in the text. Target and distractors (where possible) are paraphrased.
Meaning representation /inference	5	B2	5	1	Monologues and dialogues	Q&A about listening text. Listen to monologues and conversations to identify a speaker's attitude, opinion or intention. The information targeted will require the integration of propositions across the input text to identify the correct answer.	Three-option multiple choice. Both target and distractors are (where possible) paraphrased, and distractors refer to important information and concepts in the text that are not possible answers to the question.

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H 9 7 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

Structure of the teens speaking component

Part	Skill focus	Level	Task description	Channel of input/ prompts	Time to plan	Time for response	Rating criteria
Test: Speaking (25 items)							
1	Giving personal information	A1/A2	Candidate responds to three questions on personal topics. The candidate records their response before the next question is presented.	Questions presented in both written and oral form (pre-recorded). Questions presented in a sequence (e.g. Q2 is presented after the response to Q1).	No	30 seconds to respond to each question.	Separate task-based holistic scales are used for each task. Performance descriptors describe the expected performance at each score band. The following aspects of performance are addressed:
2	Describing, expressing opinions, providing reasons and explanations	B1	The candidate responds to three questions. The first asks the candidate to describe a photograph. The next two are on a concrete and familiar topic related to the photo.	1) Questions presented in both written and oral form (pre-recorded). Questions presented in a sequence (e.g. Q2 is presented after the response to Q1). 2) A single photo of a scene related to the topic and familiar to A2/B1 candidates on screen.	No	45 seconds to respond to each question.	1) grammatical range and accuracy 2) lexical range and accuracy
3	Describing, comparing and contrasting, providing reasons and explanations	B1	The candidate responds to two questions/prompts and is asked to describe, contrast and compare two photographs on a topic familiar to B1 candidates. The candidate gives opinions and provides reasons and explanations.	1) Questions presented in both written and oral form (pre-recorded). Questions presented in a sequence (e.g. Q2 is presented after the response to Q1). 2) Two photographs showing different aspects of a topic are presented on screen.	No	45 seconds to respond to each question.	3) pronunciation 4) fluency 5) cohesion and coherence.

Structure of the teens speaking component (continued)

Part	Skill focus	Level	Task description	Channel of input/ prompts	Time to plan	Time for response	Rating criteria
Test: Speaking (25 items)							
4	Integrating ideas on a topic into a long-turn presentation. Giving and justifying opinions, advantages and disadvantages	B2	The candidate plans a longer presentation integrating information given to them and adding their own opinion/ knowledge of the subject.	The candidate is presented with a poster which they are told they have prepared and must present to their class.	90 seconds	Two minutes for the entire response.	Separate task-based holistic scales are used for each task. Performance descriptors describe the expected performance at each score band. The following aspects of performance are addressed: <ul style="list-style-type: none"> 1) grammatical range and accuracy 2) lexical range and accuracy 3) pronunciation 4) fluency 5) cohesion and coherence.

Structure of the teens writing component

Part	Skill focus	Level	Task description	Channel of input/prompts	Expected	Rating criteria
Test: Writing						
1	Writing at the word level. Simple personal information on a form	A1	The candidate completes a form by filling in some basic personal information. All responses are at the word/phrase level, such as name, birthdate, etc.	Form with nine clearly marked categories (name, date of birth, etc.). There are nine gaps in the form to be filled.	Nine short gaps filled by one-/two-word responses	Separate task-based holistic scales are used for each task. Performance descriptors describe the expected performance at each score band. The following aspects of performance are addressed (not all aspects are assessed for each task):
2	Short written description of concrete, personal information at the sentence level	A2	The candidate continues filling in information on a form. The task-setting and topic are related to the same purpose as the form used in part one. The candidate must write a short response using sentence-level writing to provide personal information in response to a single written question.	Written. The rubric presents the context, followed by a short question asking for information from the candidate related to the context.	20–30 words	
3	Interactive writing. Responding to a series of written questions with short paragraph-level responses	B1	The candidate responds interactively to three questions. Each response requires a short paragraph-level response. The questions are presented as if the candidate is writing on an internet forum or social-network site. The task setting and topic are related to the same purpose/activity used in parts one and two.	Written. The rubric presents the context (discussion forum, social media, etc.). Each question is displayed in a sequence following the completion of the response to the previous question.	30–40 words in response to each question	1) task completion 2) grammatical range and accuracy 3) lexical range and accuracy 4) cohesion and coherence 5) punctuation and spelling.
4	Continuous paragraph-level essay writing	B2	The candidate writes an argumentative essay on a topical issue the candidate is likely to encounter in public or educational domains.	Written. The rubric presents the context in the form of an advert giving basic information about an essay competition.	220–250 words	

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W 1 0 0 6 B E N G L I S H ? I M P A C T B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H S 8 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

APPENDIX B: FINAL CFA MEASUREMENT MODEL ESTIMATES

	Estimate	S.E.	Est./S.E.	Two-tailed p-value
INTOR BY				
IO1	1.000	0.000	999.000	999.000
IO2	1.740	0.105	16.649	0.000
IO3	1.901	0.124	15.289	0.000
IO4	1.474	0.102	14.515	0.000
IDEAL BY				
I1	1.000	0.000	999.000	999.000
I2	1.035	0.024	43.279	0.000
I3	1.007	0.023	44.103	0.000
I4	1.094	0.025	43.252	0.000
PARENTS BY				
PAR1	1.000	0.000	999.000	999.000
PAR2	1.204	0.055	22.060	0.000
PAR3	1.160	0.053	21.800	0.000
PAR4	1.147	0.059	19.517	0.000
SELF BY				
SELF1	1.000	0.000	999.000	999.000
SELF2	1.019	0.020	50.313	0.000
SELF3	1.096	0.022	49.309	0.000
SELF4	1.125	0.022	51.201	0.000
EXPERIEN BY				
EX1	1.000	0.000	999.000	999.000
EX2	0.917	0.027	33.374	0.000
EX3	0.981	0.023	41.935	0.000
EX4	1.189	0.027	43.506	0.000
INSTRU BY				
INSTRU1	1.000	0.000	999.000	999.000
INSTRU2	1.018	0.033	31.079	0.000
INSTRU3	1.100	0.047	23.596	0.000
INSTRU4	0.788	0.049	16.081	0.000

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B :: 1 6 B = 6 P W 8 F W ? 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 1 0 1 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

	Estimate	S.E.	Est./S.E.	Two-tailed p-value
MOTBEH BY				
MB1	1.000	0.000	999.000	999.000
MB2	0.974	0.021	45.503	0.000
MB3	1.026	0.017	59.304	0.000
MB4	0.751	0.024	30.738	0.000
UGHT BY				
O1	1.000	0.000	999.000	999.000
O2	1.156	0.056	20.682	0.000
O3	1.168	0.052	22.278	0.000
O4	1.043	0.038	27.215	0.000
IDEAL WITH				
INTOR	0.241	0.018	13.039	0.000
PARENTS WITH				
INTOR	0.119	0.011	10.466	0.000
IDEAL	0.212	0.018	11.499	0.000
SELF WITH				
INTOR	0.181	0.017	10.922	0.000
IDEAL	0.700	0.032	21.739	0.000
PARENTS	0.181	0.020	8.944	0.000
EXPERIEN WITH				
INTOR	0.219	0.017	12.582	0.000
IDEAL	0.545	0.028	19.207	0.000
PARENTS	0.156	0.018	8.822	0.000
SELF	0.614	0.031	19.670	0.000
INSTRU WITH				
INTOR	0.190	0.016	11.951	0.000
IDEAL	0.403	0.024	16.524	0.000
PARENTS	0.326	0.023	14.418	0.000
SELF	0.386	0.025	15.324	0.000
EXPERIENS	0.347	0.025	14.119	0.000

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W 1 0 2 6 B E N G L I S H ? I M P A C T B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 H S 8 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

	Estimate	S.E.	Est./S.E.	Two-tailed p-value
MOTBEH WITH				
INTOR	0.163	0.014	11.298	0.000
IDEAL	0.401	0.025	16.013	0.000
PARENTS	0.191	0.019	10.059	0.000
SELF	0.468	0.030	15.767	0.000
EXPERIENS	0.588	0.027	21.889	0.000
INSTRU	0.349	0.023	15.239	0.000
UGHT WITH				
INTOR	0.096	0.011	8.523	0.000
IDEAL	0.148	0.021	7.177	0.000
PARENTS	0.245	0.023	10.791	0.000
SELF	0.159	0.023	6.822	0.000
EXPERIENS	0.176	0.021	8.381	0.000
INSTRU	0.282	0.022	12.636	0.000
MOTBEH	0.203	0.022	9.172	0.000
INSTRU2 WITH				
INSTRU1	0.316	0.026	12.2270.000	
IO3 WITH				
IO4	0.260	0.027	9.490	0.000
O1 WITH				
O4	0.227	0.034	6.6170.000	
EX1 WITH				
EX3	0.105	0.015	7.030	0.000
Intercepts				
IO1	5.554	0.018	311.111	0.000
IO2	5.398	0.020	266.176	0.000
IO3	4.729	0.026	179.420	0.000
IO4	4.818	0.024	197.208	0.000
PAR1	4.520	0.028	159.059	0.000
PAR2	5.011	0.026	195.581	0.000
PAR3	5.149	0.023	225.080	0.000
PAR4	4.719	0.027	172.118	0.000
SELF1	4.015	0.030	134.816	0.000

A :: F N O A G // X C K P W ? 9 B E // 1 7 B = 6 > ? 9 B E B 1 7 B = Z V 2 8 T Z Q U - < 2 5 3 5 % A
 N O A G W B N 7 = ? 9 W N O A G W B :: E = ? 9 W B 8 F W ? 9 B E B 1 7 B 9 8 1 % > ? 9 B E B 1 7 ?
 ? // G W B S P A I N , M A D R I D 9 > ? // G W B :: 1 7 B = 6 P W 8 F W ? 4 > J H Y 4 1 0 3 3 Q J H
 G % 2 A X G M L 1 9 8 1 ^ 9 # D R M < L % A 9 M 5 < Z V 2 8 T I Q U < 2 5 3 5 % A X G % L 1 U M 8
 F W B :: 1 7 B = 6 P W ? 9 B E B 1 7 O A G W B :: 7 = ? 9 = W 8 F W 8 F W ? 9 B E B W 8 F W 8 F W ?

	Estimate	S.E.	Est./S.E.	Two-tailed p-value
SELF2	3.804	0.030	128.039	0.000
SELF3	4.095	0.032	126.266	0.000
SELF4	3.943	0.032	122.872	0.000
O1	3.759	0.030	124.235	0.000
O2	3.229	0.030	107.941	0.000
O3	3.774	0.029	130.562	0.000
O4	3.523	0.028	123.876	0.000
EX1	4.523	0.026	172.929	0.000
EX2	3.602	0.025	145.104	0.000
EX3	4.453	0.026	174.571	0.000
EX4	3.976	0.027	147.355	0.000
MB1	4.165	0.025	169.583	0.000
MB2	3.909	0.026	148.637	0.000
MB3	4.142	0.025	167.954	0.000
MB4	3.343	0.025	133.498	0.000
INSTRU1	5.027	0.026	194.280	0.000
INSTRU2	4.804	0.026	184.431	0.000
INSTRU3	4.922	0.025	193.572	0.000
INSTRU4	3.836	0.030	128.717	0.000
I1	4.547	0.026	174.630	0.000
I2	4.770	0.026	180.710	0.000
I3	4.920	0.025	195.280	0.000
I4	4.647	0.028	167.825	0.000
Variances				
INTOR	0.141	0.017	8.445	0.000
IDEAL	0.823	0.036	22.568	0.000
PARENTS	0.444	0.038	11.596	0.000
SELF	1.135	0.047	23.937	0.000
EXPERIENS	0.749	0.039	19.402	0.000
INSTRU	0.507	0.036	14.271	0.000
MOTBEH	0.853	0.033	25.514	0.000
OUGHT	0.620	0.048	12.962	0.000

	Estimate	S.E.	Est./S.E.	Two-tailed <i>p</i> -value
Residual Variances				
IO1	0.424	0.022	19.574	0.000
IO2	0.301	0.018	16.662	0.000
IO3	0.721	0.034	21.373	0.000
IO4	0.751	0.032	23.219	0.000
PAR1	0.988	0.037	26.900	0.000
PAR2	0.520	0.028	18.819	0.000
PAR3	0.331	0.021	15.865	0.000
PAR4	0.749	0.036	20.573	0.000
SELF1	0.437	0.020	21.457	0.000
SELF2	0.387	0.021	18.643	0.000
SELF3	0.501	0.030	16.946	0.000
SELF4	0.390	0.018	21.129	0.000
O1	1.003	0.044	22.590	0.000
O2	0.757	0.048	15.888	0.000
O3	0.635	0.036	17.723	0.000
O4	0.759	0.040	18.918	0.000
EX1	0.464	0.019	25.002	0.000
EX2	0.463	0.020	23.428	0.000
EX3	0.434	0.020	21.787	0.000
EX4	0.233	0.015	15.577	0.000
MB1	0.216	0.015	14.604	0.000
MB2	0.417	0.019	21.906	0.000
MB3	0.179	0.013	14.045	0.000
MB4	0.630	0.028	22.280	0.000
INSTRU1	0.680	0.038	17.998	0.000
INSTRU2	0.678	0.032	20.870	0.000
INSTRU3	0.533	0.027	19.939	0.000
INSTRU4	1.260	0.043	29.613	0.000
I1	0.379	0.017	22.767	0.000
I2	0.354	0.019	18.702	0.000
I3	0.292	0.020	14.425	0.000
I4	0.375	0.021	17.909	0.000

A::FNOAG//XCKPW?9BE//17B=6>?BEB17B=ZV28TZQU-<2535%ANOAGWBN7=?9WNOAGWB::E=?9WB8FW?9BEB17B981%>?9BEB17?//GWBSPAIN,MADRID9>?//GWB::17B=6PW8FW?4>JHY41053QJHG%2AXGML1981^9#DRM<L%A9M5<ZV28TIQU<2535%AXG%L1UM8FWB::17B=6PW?9BEB17OAGWB::7=?9=W8FW8FW?9BEBW8FW8FW?



