BRITISH COUNCIL

ENGLISH IMPACT

AN EVALUATION OF ENGLISH LANGUAGE CAPABILTY

UZBEKISTAN

MARCH 2023

KAREN DUNN
JANINA IWANIEC

FOREWORDS

Denise Waddingham

British Council Country Director, Uzbekistan

The British Council has been working in Uzbekistan since 1996, supporting the development of education and particularly the development of English language competence across the country. The past five or six years since 2017 have seen rapid transformation in Uzbekistan, with a growing need for foreign language knowledge, and a massive increase in the demand for English.

The British Council's work in English includes reform of the national pre-service teacher training (PRESETT) curriculum; mapping of the professional development journey of an English teacher in public education and development of major principles of Continuing Professional Development (CPD); modern English Curriculum and assessment consultancies and training for major stakeholders; cutting-edge digital resources and products and delivery of modern, communicative, four-skills examinations for teachers and learners of English.

Our current flagship programme with the Ministry of Pre-School and School Education of the Republic of Uzbekistan is called Future English. It aims to understand and address the perceived issue of the low level of English skills of school leavers and to develop solutions to bring about sustainable change.

Based on the extensive research into English teaching skills in public schools conducted by the British Council in 2020, we developed the Online Teacher Community (OTC) programme that brings together teachers from different countries in the world and offers a number of tailored courses around teaching skills, classroom research, assessment and English for teachers. Annually around 4000 English teachers from Uzbekistan join our OTC.

English Impact is one of the research strands of Future English and was designed to build a multi-dimensional profile of the English capability of school students that will help to demonstrate the impact policy decisions can have on the development of learning systems: curriculum, delivery and assessment.

At the British Council, this project is particularly important for us given our commitment to enable more widespread and better-quality training, teaching, learning and assessment of English worldwide. Although completion of the project was delayed by the impact of the Covid-19 pandemic, we hope that the results of the English Impact report will provide the government and other interested stakeholders with essential data and evidence to inform future policies and activities.

Barry O'Sullivan

Head of Assessment Research & Development, British Council

I've had the privilege of introducing three previous English Impact reports and I'm delighted to introduce this current report which focuses on Uzbekistan. The three reports published to date focused on two regions, Madrid (2017) and Bogota (2018), and one country Sri Lanka (2018); they have added considerably to our knowledge of the situation regarding the English language proficiency of young leaners in those places. The experience gained in designing, delivering and interpreting the results from those three studies has contributed significantly to the Uzbekistan English Impact project and to the report you are reading now. Comparisons have been made between the English Impact studies and the OECD's Programme for International Student Assessment (PISA), in particular the upcoming foreign language assessment. I think it is important to note that, while the latter has the potential to offer governments a uniform measure whereby inter-country or region comparisons might be drawn, the fact that the English Impact studies are designed in tandem with the local Ministry of Education means the findings are more likely to be of specific value to that Ministry. A good example of this is in the selection of the stratification variables which are essentially generic within the PISA system but context-focused for the English Impact studies (e.g., see Table 4.6).

As with the earlier studies, the project team used the British Council's Aptis for Teens English language test as the principal measurement instrument. This allowed for detailed comparison across the population on the four skills (listening, reading speaking and writing) as well as on knowledge of the language itself (grammar and vocabulary). In addition to this, we again used the highly-rated motivation questionnaire designed by Dr Janina Iwaniec from the University of Bath, UK. Dr Iwaniec also undertook the analysis of the questionnaire data, while additional analyses were undertaken by Dr Karen Dunn from the British Council, UK and by Dr Martin Murphy and his team at the Australian Council for Educational Research (ACER). ACER also worked with the project team to develop the complex sampling strategy that underpins the whole project. While conceptualising and designing a project of this magnitude is clearly important, of no lesser importance is the work of the teams on the ground in Uzbekistan who worked tirelessly to achieve the demanding response rates to ensure a representative sample.

The previous English Impact reports have had a significant impact on regional and national education policies. The robustness of these studies have contributed to this impact. I fully expect that this study will have the same level of impact on English, and other, language policy in Uzbekistan.

Martin Murphy

Director of ACER Sampling, ACER

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 policymakers 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 the 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 Uzbekistan and the British Council teams 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 Uzbekistan'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 Uzbekistan and 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 Uzbekistan would not have been possible:

Ministry of Pre-school and School Education of the Republic of Uzbekistan (at that time Ministry of Public Education) and the below contributors in particular:

- **Laziz Khujakulov**, Head of Strategic Development Department
- **Usman Sharifkhojaev**, First Deputy Minister of Pre-school and School Education

Colleagues in **British Council Uzbekistan** for their implementation of English Impact:

- **Jamilya Gulyamova**, Deputy director
- **Saida Irgasheva**, Projects coordinator
- **Sadakat Khojaeva**, Projects Co-ordinator
- **Olga Kim**, English Projects Manager
- **Andrey Maksakov**, Head of Exams Business
- **Komila Muminova**, Test Day Delivery Co-ordinator
- **Margarita Zinurova**, Planning and Resourcing Coordinator

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

- **Bethany Davies**, Research Fellow
- **Jorge Fallas**, Research Fellow
- **Sladana Krstic**, Senior Research Fellow
- **Brendan McGinley**, Research Fellow
- **Abigail Middel**, Research Officer
- **Martin Murphy**, Director of ACER Sampling

Motivation questionnaire design and analysis:

- **Janina Iwaniec**, Senior Lecturer, Department of Education, University of Bath

Uzbekistan Educational Context Information

- **Andy Keedwell**, Training Consultant

British Council project team:

- **Tatjana Slijepcevic**, Project Manager Hub Wider Europe
- **Joseph Field,** FE Programme Senior Responsible Owner
- **Addeh Hovassapian**, FE Programme Senior Project Manager
- **Maja Mandekic**, (Formerly) British Council Regional Portfolio Manager
- **Marijana Petrovic**, Research Coordinator (British Council appointed)

Members of the **British Council Assessment Research Group**:

- **Karen Dunn**, Senior Researcher
- **Jan Langeslag**, Emerging Researcher
- **Barry O'Sullivan**, Head of Assessment Research & Development

1. EXECUTIVE SUMMARY

English Impact provides robust policy-relevant data evidencing English language capability in Uzbekistan. The concept of capability is characterised by an understanding of both current achievement and future opportunity in English language learning, captured within a multi-dimensional profile.

English Impact research combines information about the current English language proficiency of a targeted sample of students from the publicly funded school sector, with in-depth analysis of the language learning experiences in and outside the classroom of these students, plus their language learning motivations. To ensure that the data collected by English Impact Uzbekistan can be used to inform and support education system and policy development, the British Council worked in collaboration with the Ministry of Education tailoring the research to meet local needs. This research is underpinned by the British Council's Royal Charter and charitable objective to develop a wider knowledge of the English language and it 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. The research was carried out by the British Council, with contributions from the Australian Council for Educational Research (ACER) and the University of Bath.

English Impact in Uzbekistan

- English Impact employs a two-stage cluster sample design used by other recognised large-scale international surveys, sampling schools at the first stage and students at the second stage.
- 150 government-funded schools and 1,446 students were sampled for English Impact, with 121 schools and 1,331 students participating following exclusions, student withdrawal from school or absence.
- Students were sampled from Grade 10 in compulsory secondary education. This grade represents 10 years of schooling, counting from the first year of International Standard Classification of Education (ISCED) Level 1, with a mean age at the time of testing of 16.3 years.
- Students sampled were studying English as part of their studies at this grade level. A minimum of 90 minutes of formal English study per week as part of the school program was required for eligibility in the target population.
- Students completed the British Council's Aptis for Teens English Language assessment, testing reading, writing, speaking, listening, and grammar and vocabulary.
- A questionnaire comprising 53 items, delivered in Uzbek, gathered opinions and information from students on their schooling and language learning experience, including engagement with digital media, and their language learning motivation.

Key findings

- Overall school and student participation in Uzbekistan met the English Impact international participation standard of at least 85 per cent of sampled students in 85 per cent of sampled schools. As such, the information given by the study is highly relevant for informing policy decisions and can act as a robust baseline against which future comparisons can be drawn.
- 75% of the students indicated that they intend to continue education to tertiary level and 10% plan to take on some form of vocational training.
- 60% of Uzbek students started learning English in Grade 5, a further 18% started prior to grade 5 and 19% whilst in grades 7-9.
- Within the English language classroom 49% of students reported to have the chance to practice speaking English in class, and 51% of students reported to be given the opportunity

to practice English in pair and group work in class regularly, a lot or all the time.
- Just under 50% reported to study English in lessons outside of school.
- Digital media reported to be accessed regularly in English included computer games (44% of participants), social media (almost 30%) and other online content (27%).
- Over 90% of students report that they regularly use digital resources for English learning either in the classroom, or at home, or both.
- 68% of the sampled students from Uzbekistan achieved at A1 level in their overall English language test performance on the Common European Framework of Reference for Languages (CEFR); 21% achieved at A2 level overall.
- Students performed overall better in the receptive language skills (listening and reading) than the productive language skills (speaking and writing).
- For listening skill performance, 17% of participating students achieved at B1 level, whilst a high proportion, 67% achieved at A2 level. For reading 53% of students achieved A2 and 29% A1 level.
- In writing 27% of students achieved at A1 level and 11% at A2, and for speaking 16% at A1 and 8% at A2 level.
- A positive and significant difference between urban and rural schools' performance was identified in all language skill areas, with urban schools outperforming rural schools.
- No significant differences were found between female and male students in English language test performance.
- Findings indicate that male and female students perceive English equally as important for their professional career and feel similar levels of pressure and encouragement from their environment, including their parents, to study English.
- Motivational traits reflecting family and social expectations are least closely related to motivated learning behaviour. This is in contrast with more internalised motivational traits which all have much stronger relationships with motivated learning behaviour. In other words, effort invested in language learning is more likely to be evident in teenagers who understand and have, to some extent, internalised the value of English language learning.
- Although relationships between language learning motivation and proficiency is only a small part of the full picture of the language learner, a positive relationship between English language proficiency and motivation levels emerges from the study. This applies particularly to the traits reflecting internalised motivations.

2. INTRODUCTION

The English Impact research methodology was designed in 2017 to provide a robust evaluation of the English language capability of teenage language learners at the heart of where government policy makes an impact: in publicly funded school classrooms (Shepherd and Ainsworth, 2017, 2018a, 2018b). The concept of capability is characterised by an understanding of both current achievement and future opportunity in English language learning, captured within a multi-dimensional profile. Findings from the English Impact Uzbekistan study incorporate empirical evidence of English language proficiency of language learners at schools across the country, alongside information about these students' language learning experiences and motivations. The aim of garnering this unique insight is to aid in the diagnosis of the impact policy decisions can have on the development of learning systems: curriculum, delivery and assessment.

English Impact Uzbekistan

English Impact in Uzbekistan was instigated in 2019, when the British Council and the Ministry of Public Education of the Republic of Uzbekistan signed a letter of agreement on running the Future English project in Uzbekistan. The programme was initiated to address the reported and perceived problem of low level of English skills of school leavers who are entering work or tertiary education. Two project strands focused on conducting the research to generate reliable and valid data to identify and inform solutions. English Impact provides empirical insight into the English language capability of school students, whilst the other strand of research provided a teacher-focussed needs analysis to help shape future policies aimed at improving and developing effective national CPD (continuous professional development) systems. The ultimate aim of the Future English project is to improve teaching skills and practices in the classroom which will lead to increased learning outcomes and ensure that students are better equipped to advance in higher education, become internationally mobile and increase their employability prospects in the labour market[1]. The English Impact strand of the research was unfortunately delayed owing to the Covid-19 pandemic, meaning that the instigation of the English Impact survey was delayed until the latter part of 2021. It was only at this point that the situation around Covid-19 in Uzbekistan became more stable and the data collection and testing could be safely run in the sampled schools.

Empirical insights for policy development

English language learning now plays a significant role in many national and regional education systems, with 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. Considerably harder to find are 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. This global best practice in research, together with experience of data collection, is emulated within the design of the English Impact methodology that will be detailed in the following chapters. Much can be learnt from the decades of experience in designing the processes to sample and implement large-scale research of this kind.

1. Further information about the Future English project and additional strands can be found at: https://www.britishcouncil.uz/en/programmes/education/future-english/future-english-uzbekistan

Whilst 2025 will see the roll-out of the English language component in PISA (Marconi et al., 2020), providing an additional facet to this survey and a source for international comparisons, the insights from English Impact directly feed into this picture, providing a dedicated focus on English language capability which involves a rich and nuanced profile of grade 10 teenagers with respect to their language learning experiences and motivations, alongside proficiency levels across four skills.

Following its inception in 2017, English Impact has been successfully implemented in five countries globally, the findings have been used to inform policy decisions and further research. In Madrid, for example, the research provided insights into the bilingual education system in this region of Spain and engendered further investigations of a more qualitative nature into the CLIL (Content and Language Integrated Learning) classroom and teaching practices (Iwaniec and Halbach, 2021).

English language capability

The theoretical basis used to define English language capability is derived from an adaptation of Amartya Sen's capabilities approach. 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 become 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', Saito (2003) links this capability approach to education and the opportunity to achieve. English language capability can therefore be described in terms of the level of achievement, or proficiency, reached by a defined population, plus 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. The evaluation of English language capability, reflecting Sen's capabilities approach, is therefore not limited to the measurement of English language proficiency as captured by a test. Other data were captured and combined to provide a broader context to our analysis: educational context, language learning environment, language proficiency, and language learning motivations. The presentation of these supporting data is intended to provide a broader context for understanding and interpreting students' assessment outcomes, to elaborate the functioning of current policy, and provide a solid empirical platform for discussions surrounding English language policy in Uzbekistan moving forward.

Research aims

The research aims outlined and investigated were:

- to evaluate the English language capability of grade 10 students studying English at public schools in Uzbekistan
- to compare the capability of students attending urban and rural schools, and schools across the 14 regions in Uzbekistan
- to gain insights into comparative levels of language learning motivation and to understand the relationship between motivation and English language 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.

3. RESEARCH METHODOLOGY

The English Impact research methodology was designed to provide the most credible evaluation of the English language capability in Uzbekistan. 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 potential to succeed. Fundamental to the evaluation of English language capability within a national or regional education system is an appropriate sampling methodology employed to accurately reflect the population of interest and supply sufficiently precise estimates from the survey. The research methodology employed in this English Impact study involves three central components discussed in turn below:

- 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), principally the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), as well as surveys of 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 Uzbekistan 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 employed is given below.

I. Preparation

- define the comparison unit (see further details in highlight box)
- identify exclusions
- determine stratification variables
- obtain database of schools and agree access
- agree the sample design.

II. School sampling

- select the school sample.

III. 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.

IV. Data tracking

- track school participation
- track student participation.

The British Council team in Uzbekistan worked directly with the ministry of public education and regional administrative centres and the ACER research team to gather the school and student level data needed to complete the sampling process.

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. This provides a clear delineation of the geographical area to which the findings of the survey can be generalised. This does not always refer to a nation, since 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.

Like much of the English Impact sample methodology, precedents established via PISA's use of adjudicated regions and TIMSS's use of benchmarking entities alongside units of comparison have informed the comparison unit policy implemented throughout the English Impact research. Close adherence to the participation standards and population definition, in combination with concisely described and internationally recognisable units of comparison, ensures that the extent of generalisations possible from the findings can be clearly elucidated.

Participation standards

The English Impact research was guided by an established set of participation standards drawn from those used within established international surveys. Use of these standards enabled precision around the major estimates of the research, namely English language capability, and maximised 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 the '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 5% 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 Uzbekistan 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 Uzbekistan from the comparison unit.

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

Standard 1.8 The English Impact Uzbekistan student response rate is at least 85% 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.

Target population definition

The international target population is defined as follows:

Students within the comparison unit enrolled in the grade that represents ten years of schooling, counting from the first year of ISCED Level 1, providing the mean age at the time of testing is at least 15 years and 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 participating in English Impact studies. 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. Some further details on the definition are explained below.

- The International Standard Classification of Education (ISCED)

UNESCO's International Standard Classification of Education (ISCED) is an internationally recognised classification of the levels of schooling across countries, ranging from pre-primary education (ISECD 0) all the way 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 population 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 Uzbek students meeting this definition and studying at least 90 minutes of formal English learning per week, rather than for the entire student population of Uzbekistan.

Precision of estimates

The primary basis for determining sample sizes is the desired precision of major observations from the survey. It is common practice to present this measurement in the form of standard errors and/or confidence intervals around survey estimates, as will be provided in the presentation of English Impact research outcomes. These sample size recommendations are followed for each comparison unit included in an English Impact survey:

- 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 TIMMS 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. For this level of precision, these surveys aim for countries 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 score points. This standard error means a 95% confidence interval of ±7.0 score 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% of the percentage estimate. This means that the confidence interval around population percentage estimates should be less than ±3.5%

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-level exclusions may include schools that are very remote or very small
- student-level exclusions include students with either functional or intellectual disabilities that prevent them from taking part in the assessment, fitting 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 5%.

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 Uzbekistan a pioneering digital method of delivery was developed. Every English language assessment and student questionnaire were completed by students via a completely offline enabled laptop computer supplied by the British Council. Whilst other large-scale assessments such as PISA and TIMSS have made initial steps towards computer-based assessment; the 2017 English Impact was a pioneering completion of a large-scale assessment using 100 percent computer-based delivery and the delivery in Uzbekistan in 2021 followed suite.

Data were collected via two applications (apps) on each laptop in fully invigilated conditions. 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 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 British Council's Aptis for Teens test (O'Sullivan, Dunlea, Spiby, Westbrook & Dunn, 2020). This test assesses four language domain-specific skills: reading, writing, speaking and listening, plus a core component comprising grammar and vocabulary.

The Aptis test system

Aptis is a computer-based test of general English proficiency and currently 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 13–17 year olds. An important feature of the tests developed within the Aptis test system 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 different 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. An overview of the CEFR global scales is given in Appendix A.

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 which 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 Uzbekistan, all five components were taken.

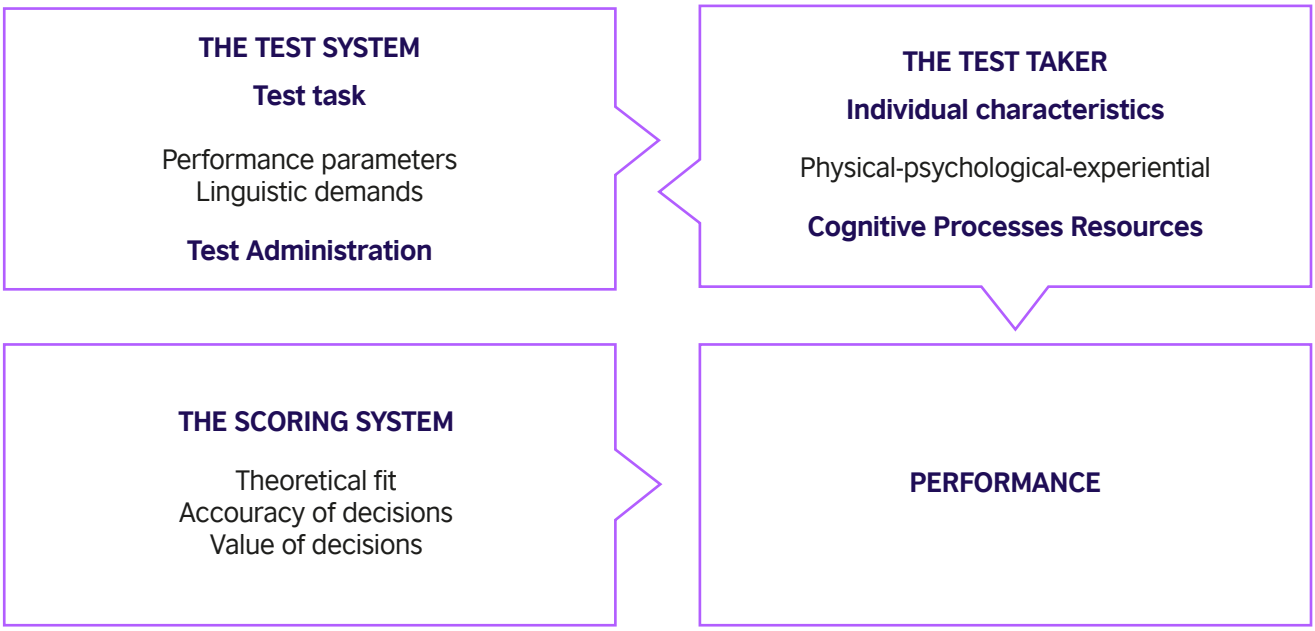
Theoretical model underpinning the Aptis test system

The theoretical model of test development and validation which 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 which 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 3.1, taken from O'Sullivan (2015), demonstrates how the three elements feed into the test taker's performance.

Figure 3.1: 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 perform to demanding technical performance criteria. The Assessment Research Awards and Grants (ARAGs) scheme actively funds research into the tests from leading international researchers. An impressive body of published documentation, covering an extensive and diverse range of validation projects, contributes important evidence to the validity argument supporting the uses of the Aptis test system. Access to the reports is available on the following webpages:

- https://www.britishcouncil.org/exam/aptis/research/publications/technical
- https://www.britishcouncil.org/exam/aptis/research/publications/validation

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 different 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 et al. (2020). The description for level four is reproduced below in Table 3.1.

Table 3.1: Level of localisation for Aptis for Teens (from O'Sullivan et al., 2020)

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 Advanced, Aptis for Teens).

Aptis for Teens

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 intended audience are made 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 competencies of the age group are also taken into consideration. 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 please see Appendix B.

Aptis for Teens was designed for a specific age group, young learners aged from 13 to 17 years. 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 British Council employees, who were present in 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 core (grammar and vocabulary), 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 skills component, the core component is reported only on the 0–50 scale. 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 in English. 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 these answers are marked by trained raters. See Appendix B 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 questionnaire was delivered in Uzbek, meaning that responses did not rely on English language proficiency.

The student questionnaire comprised 53 items in four sections.

- Demographic background

Items within this section of the questionnaire include grade, gender, age, prior schooling, language spoken within the home, and country of birth.

- 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 in subject area lessons learning English, and whether students study English outside of school. There are also questions about classroom experiences and the students' perceptions of the approaches taken by their English teacher(s).

- The use of technology and learning English

Questions in this area were included to gauge and understand the language used when students interact with various types of media, both off- and online. There were also questions focusing on the students' use of digital resources in English language learning both at home and at 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.

The motivational scales (further details are given below) were measured by 32 questions addressing eight hypothesised traits of language learning motivation, based on second language acquisition theory. Including four questions for each scale increases the value of the information gathered for each area, something which is exploited in the analytic approach, also described below. Grouped together in one section of the questionnaire and in a randomised order, participants were asked to give a response to each statement using a six-point Likert scale.

The motivational dimension

In line with the aims of the English Impact project to capture information about future potential to succeed as well as students' current English language ability, the motivation questions were constructed with strong theoretical and empirical grounding to provide insights into underlying motivations and attitudes to English as an international language.

Motivation is one of the most influential of all individual differences, trumping factors such as language learning aptitude (Gardner & Lambert, 1972) in explaining gains in proficiency in certain contexts. Recently, it has been shown that motivation is more important than the age of 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 & Singleton, 2016). Motivation is also considered to be responsive to appropriate interventions (Taylor & Marsden, 2014) and can be enhanced or decreased as a result of language learning environment (Ushioda, 2009). Both 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 & 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 & Ushioda, 2009) which consists of three constructs: ideal L2 self, ought-to L2 self and language learning experience; international orientation (Yashima, 2000); and self-concept (Bong & Skaalvik, 2003). Full details are given in Table 3.2.

Table 3.2: Reference table of motivational scales and associated descriptions

Motivational scale (latent variable)	Descriptive name	Motivational scale descriptions	Question ref	Question details
Ideal L2 self (IDEAL)	Personal language goals	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, ideal L2 self 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.	I1	I imagine myself speaking English fluently.
			I2	I imagine myself comfortably reading in English on the internet.
			I3	I imagine myself easily being able to follow what others say to me in English.
			I4	I imagine myself writing online in English with ease.
Ought-to L2 self (OUGHT)	Social expectations	The ought-to L2 self is based on the external 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, pp. 105-106).	O1	I consider learning English important because the people I respect think that I should do it.
			O2	Studying English is important to me because other people will respect me more if I have knowledge of English.
			O3	Studying English is important to me because an educated person is supposed to be able to speak English.
			O4	Learning English is necessary because people surrounding me expect me to do so.

Motivational scale (latent variable)	Descriptive name	Motivational scale descriptions	Question ref	Question details
Language learning experience (EXPER)	Interest in learning English	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)	Future opportunities	Instrumentality represents motivation stemming from the practical benefits of language (Gardner & 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)	Global communication	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, p. 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 different 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.

Motivational scale (latent variable)	Descriptive name	Motivational scale descriptions	Question ref	Question details
English self-concept (SELF)	Self-confidence in English	Self-concept is 'a person's perception of himself' (Shavelson, Hubner, & Stanton, 1976) and this scale relates to self-evaluation in the students' ability to study English. The most common measurement of self-concept is Marsh's (Marsh, 1990) Academic Self-Description Questionnaire, 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)	Family expectations	Like the ought-to L2 self, parental encouragement focuses on external expectation. As the participants in English Impact are as young as 15.5, 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 & 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 practice my English as much as possible.
Motivated learning behaviour (MOTIV)	Level of effort	Motivated learning behaviour attempts to measure the behavioural component of motivation, i.e. the reported amount of effort 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 indicate a common underlying trait. No individual observed 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 several observed variables in order to give information about the levels of a hypothesised latent variable.

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, comparative male/female student groups were examined

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.

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. CFA 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 of the variation in any 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 upon 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, then the model is considered to be a good fit. For the fit indices reported here, is usually expected that the CFI and

TLI statistics be above 0.9 (or ideally 0.95), and the RMSEA below 0.05 for good model fit[2]. Where alterations are made to the initially hypothesised model, comparisons are made using the adjusted chi-square difference test (Satorra & Bentler, 2010). This is the recommended means of comparing CFA models estimated using the maximum likelihood with robust standard errors (MLR) approach[3] employed in the current analysis (Muthén & 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.

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, pp. 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 traits listed in Table 3.2, 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, 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 vary where significant differences are found. 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)[4]:

- scores from the Aptis test: overall, and for each test component
- factor scores for each participant for each motivational scale

2 See Byrne (2012, pp. 69–77) for a description of what these indices represent and the values accepted to show good fit.

3 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 & Muthén, 2011, p. 533).

4 P-values are derived from a comparison of the correlation coefficients following a Fisher's z-transformation. Calculations performed using this internet resource: http://vassarstats.net/rdiff.html

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 factor scores meanwhile were derived from the multi-group CFA described above and ascribe each questionnaire respondent with 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, he/she would have a higher factor score 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 factor scores is determined by the factor loadings estimated in the model. The factor scores themselves run on a standardised zero-centred continuous scale (i.e., between -1 and 1).

4. IMPLEMENTATION IN UZBEKISTAN

The main objective of the sample design for English Impact is to present the most accurate possible results based on the comparison unit chosen: government-funded schools in Uzbekistan. To meet the established participation standards described in Chapter 3, a precisely defined comparison unit had to be outlined and agreed. This was tailored to the Uzbekistan context; some background to the national educational context is given below, followed by a definition of the comparison unit and stratification variables.

Country profile

A landlocked country with a population of around 35 million, and classed as a lower-middle-income country, Uzbekistan is now in its fourth decade of independence from the Soviet Union. Economic growth has fluctuated since independence and is, at present, forecast for 2023 at 5.7 per cent, a slight decrease on recent years. An inflation rate of around 12 per cent remains a concern, although this appears to be stabilising. World Bank data gives GDP per capita as USD$1,983.10 for 2021 with state budget projections for a current figure as a slightly higher USD$2,100. Rates of unemployment have consistently averaged between 6 and 7 per cent over recent years and currently stand at 6.9 per cent. Youth unemployment rates may be double that figure, and it is estimated that a further 20 per cent of the population are underemployed.

The country is a major producer of cotton. Exports have been historically limited by overseas restrictions. While other crops such as wheat have grown in importance, cotton production has been encouraged by a relaxation of bans by countries such as the United States in 2022. Extraction of minerals (including gold, silver, copper and uranium) and the oil and gas sectors are also economic mainstays. There are significant variations in relative economic prosperity across the twelve provinces, one autonomous republic (Karakalpakstan) and one stand-alone city (Tashkent, the capital) which comprise the country.

Figure 4.1: Map of regions in Uzbekistan

Schools and learners

Compulsory free education extends to all children in Uzbekistan. Historically, learners completed nine years of education with four years at primary level (beginning at age seven) and five years at secondary level. Under the General Secondary Education (GSE) programme, this has been extended to an eleven-year system with a further two years either at the more academic gymnasiums/lyceums or the more vocationally-focused technical schools. Schools are still in the process of being re-integrated into the newer system, through which compulsory education will also be extended to six year olds. Systems are summarised in Table 4.1.

A private education sector was prohibited by law until 2017 when a liberalisation programme led to a steady growth of private schools (including a number of international schools in the capital, Tashkent). In comparison to the number of state schools, the number of private schools remain tiny nationally. Religious education does not figure in the state school curriculum and the number of institutions imparting any type of religious education is strictly limited by law and remains very small.

Table 4.1: Education in Uzbekistan

Age	Grade	Institution	
3 to 6 years		Preschool (GSE)	
7 to 10	1 to 4	Primary (GSE)	
11 to 16	5 to 9	Lower Secondary (GSE)	
16 to 18	10 to 11	Upper Secondary Academic Lyceum/gymnasium (also referred to as high schools)	Upper Secondary Vocational Technical School
18 and above		University: Bachelor's degree 4 to 5 years	Vocational: professional qualifications 0.5 to 2 years

Uzbekistan has a relatively young population, with 60 per cent of the population under thirty and a median age of 27 years. Eighteen per cent of the population are of school age. A total of 6,417,000 learners engage in education (compared to 6.3 million in 2021–22)[5] attending 10,522 general education institutions,[6] the vast majority being state-managed.

Figures for the number of children at school are shown in Table 4.2 below. Figures for primary and secondary enrolment are forecast to rise steadily over the next decade.[7] One significant feature of the Uzbek education system is the number of children enrolled in school, which is near universal at primary and secondary levels.[8] GER does vary from district to district, however. It is at its highest (103 per cent) in Tashkent and decreases to 93 per cent in one of the most challenged areas of the country, Karakalpakstan, although there has been substantial progress over recent years in the latter area and other similar rural areas.[9] The result is a second positive feature in that Uzbekistan has an official literacy rate of 100 per cent.[10] Gender parity is also positive (0.99 at primary/secondary level in 2021).

Uzbekistan has struggled to address the needs of SEND learners: in 2019, an average of only 1.3 per cent of SEND learners attended public schools[11] although there has been subsequent progress in this area and the need for further support has been acknowledged by relevant ministries.

5 Figures provided by relevant ministries to British Council Uzbekistan 13/3/2023

6 Accessed from: https://kun.uz/en/news/2023/01/13/number-of-schools-in-uzbekistan-increased-by-748-over-past-5-years

7 Education Sector Plan 2019–2023, Government of Uzbekistan. Available at: https://www.globalpartnership.org/sites/default/files/2019-04-gpe-esp-uzbekistan.pdf

8 CES Factbook Education System: Uzbekistan NO. 11/2021. Available at: https://ethz.ch/content/dam/ethz/special-interest/mtec/edusys-dam/documents/CES%20Factbook%20Education%20Systems_Uzbekistan.pdf

9 Accessed from: https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=UZ

10 Accessed from: https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=UZ

11 Education Sector Plan 2019-2023, Government of Uzbekistan

Table 4.2: Numbers of children attending school in Uzbekistan (preschool, primary and secondary)

	2017–2018[12]	2021–2022[13]	2022–2023[14]
Preschool	World Bank estimates at around 1,600,000		
Primary	2,391,000	2,460,100	2,544,000
Secondary Grades 5 to 9	3,732,000	2,99,720	3,064,400
Grades 10 to 11		821,400	831,500

There is, however, a sharp contrast between enrolment rates at primary/secondary level and the rate of enrolment at preschool level where substantial numbers of eligible children (a little less than 40 per cent) are not being schooled.[15] This low enrolment has been attributed to a number of factors, including lack of access, distance of available institutions from home, and the fact that many preschool institutions, unlike primary and secondary schools, charge fees which exclude the poorer sections of society. There are variations between urban and rural centres (86 per cent of children attend preschool in Tashkent compared to 41 per cent in some rural areas) and gender (in some rural areas, 46 per cent of boys attend preschool compared to 36 per cent of girls).[16] The preschool sector is also comparatively less regulated, without any systems for rigorous inspection, and quality standards vary significantly.

Secondary school is completed by 96.2 per cent of children, which compares very favourably to global averages.[17]

Intake to undergraduate degree programmes is from the academic lyceums and regulated through national examinations. In 2021, 571,500 learners were studying at higher education level, a significant increase from 183,000 in 2001.[18] Areas outside the capital, however, tend to be significantly under-represented: 35 per cent of higher education students are from Tashkent, even though the population of Tashkent is less than 6 per cent of the total for the country. [19]

Following independence, the medium of education switched from Russian to Uzbek. Russian acquired the status of a foreign language along with English and other European languages. Fewer than 7 per cent of schools now use Russian as a medium of instruction and very small percentages of schools use ethnic languages (Karakalpak, Turkmen) or the languages of neighbouring countries (Tajik, Kazakh, Kyrgyz).[20]

English as a foreign language is delivered through two classes per week at primary level, increasing to three lessons per week at higher grades.

12 Uzbekistan National Education Profile. Available at: https://www.epdc.org/node/356.html

13 Figures provided by relevant ministries to British Council Uzbekistan 13/3/2023

14 Figures provided by relevant ministries to British Council Uzbekistan 13/3/2023

15 UNICEF Education Sector Analysis 2021. Available at: https://uzbekistan.un.org/sites/default/files/2022-05/Edu%20Sit%20An_UNICEF%202022_0.pdf

16 UNICEF Education Sector Analysis 2021

17 CES Factbook Education System: Uzbekistan NO. 11/2021. Available at: https://ethz.ch/content/dam/ethz/special-interest/mtec/edusys-dam/documents/CES%20Factbook%20Education%20Systems_Uzbekistan.pdf

18 National Today 1/10/22

19 Education Sector Plan 2019-2023, Government of Uzbekistan. Available at: https://www.globalpartnership.org/sites/default/files/2019-04-gpe-esp-uzbekistan.pdf

20 Education Sector Plan 2019-2023, Government of Uzbekistan. Available at: https://www.globalpartnership.org/sites/default/files/2019-04-gpe-esp-uzbekistan.pdf

Teachers

A little over half a million teachers work in Uzbekistan's primary and lower secondary schools, around a quarter at primary level and three-quarters at lower secondary (World Bank, 2021)[21]. Around two-thirds of teachers of all subjects are female.

In a 2019 British Council/Leicester University study of English secondary school teachers, 82 per cent of respondent teachers were female.[22] Numbers of teachers at preschool, primary and secondary levels are shown in Table 4.3 and numbers of English teachers in Table 4.4.

Table 4.3: Number of teachers (all levels)

Total number of teachers: October 2022[23]	
Preschool	116,000 (other sources estimate 120,000[24])
General education (primary and secondary)	510,000
Lyceums (upper secondary)	3,300
Vocational (upper secondary)	9,900
Colleges	3,900
Technical institutions	4,900
Higher education	37,400

Table 4.4: Number of English teachers (primary, secondary and higher education)[25]

Level	Number of teachers 2023
Primary and secondary	42,828
Higher education	3,198

Pupil Teacher Ratio (PTR) is comparatively low by global standards, with variation from urban centres to rural areas where the ratio is significantly lower. PTR in Uzbek schools is summarised in Table 4.5.

However, issues such as teacher shortages mean that this low PTR does not necessarily imply corresponding small class sizes.

21 https://tradingeconomics.com/uzbekistan/primary-education-teachers-wb-data.html

22 Secondary Language Teaching in Uzbekistan, British Council, 2022

23 stat.uz Available at: https://stat.uz › Home › Press Service › Committee News

24 Infinity Group Education Market Intelligence Report/Finspect Advisory. Available at: https://finspect.uz/wp-content/uploads/2020/04/infinity-Group-Education-market-intelligence-report-Uzbekistan.pdf

25 Figures provided by relevant ministries to British Council Uzbekistan 13/3/2023

Table 4.5: Variations in Pupil Teacher Ratio (PTR)

	National average	Highest PTR	Lowest PTR
PTR	12.2:1	18:1 (Tashkent)	9:1 (Karakalpakstan, Navoi)

The extent to which teachers are qualified has been expressed as a significant concern in various reports, including the Government of Uzbekistan Education Sector Plan. At the preschool level, only 21.9 per cent of teachers have engaged in higher education, a figure which rises to 81.7 per cent at secondary level.[26] There are also significant differences in the number of staff qualified at higher education level between urban and rural areas. The British Council/Leicester University study also found high rates of inexperience among English teachers with 66 per cent of teachers having less than ten years' experience although, for historical reasons, levels of inexperience may be higher for this specific subject.[27]

Educational reform

Supervision of education in Uzbekistan is vested in the Ministry of Public Education (MPE) and the Ministry of Higher and Secondary Specialised Education (MHSSE), as well as the Ministry of Preschool Education, which was created in 2017 to address some of the issues described above.

Education is financed through a combination of national and regional budgeting. Since independence, education has consistently been seen as a major priority for government spending, reaching a high of 7 per cent of GDP between 2013 and 2016. This figure reduced to 4.8 per cent for 2021 but it is forecast to rise for 2023. Over the last ten years, spending on education has remained at between 25 to 30 per cent, a higher proportion than a number of Uzbekistan's neighbours.

The government of Uzbekistan has emphasised its commitment to an ambitious programme of comprehensive, systematic reform. This year, 2023, has been termed 'the Year of Quality Education' and the government has acknowledged the need to reduce the level of rote learning and to support the professional development of teachers.
A number of interventions are in the process of initiation or are underway. These include:

- efforts to improve school infrastructure including the renovation of 3,000 schools[28] and improved access to IT and digital learning
- provision of free school meals across the country[29]
- widening access to preschools for all children and the quality of preschool education through the Second Education Sector Plan
- further development and full implementation of the Uzbekistan National Curriculum
- initiatives through the Agency for Presidential Educational Institutions to develop new specialised schools for the gifted
- textbook revision for subjects including Uzbek, arts and mathematics and associated activities in collaboration with USAID through the Uzbekistan Education for Excellence programme (UEEP)
- a review of teacher professional development in collaboration with British universities and the British Council and subsequent action.

26 Education Sector Plan 2019-2023, Government of Uzbekistan. Available at: https://www.globalpartnership.org/sites/default/files/2019-04-gpe-esp-uzbekistan.pdf

27 Secondary Language Teaching in Uzbekistan, British Council, 2022

28 Accessed from: http://uzdaily.com/en/post/73012

29 [21]Accessed from: https://www.uzdaily.uz/en/post/79540#:~:text=Tashkent%2C%20Uzbekistan%20(UzDaily.com,in%20all%20schools%20in%20Uzbekistan.

The comparison unit

The comparison unit included Uzbek publicly-funded schools which incorporated an English language programme for grade 10 students. Implicit stratification variables were chosen to ensure a proportional sample allocation across the implicit strata. The stratification variables applied to the Uzbekistan sample frame are outlined in Table 4.6.

Table 4.6: Stratification variables

Stratification variable name	Variable labels
Administrative region	Andijan/ Bukhoro/ Fergana/ Jizzah/ Karakalpak AR/ Kashkadaryo/ Khorazm/ Namangan/ Navoiy/ Samarkand/ Sirdaryo/ Surhandaryo/ Tashkent City/ Tashkent Region
Location	Urban/Rural
Medium of instruction	English specialised (SES)/ English non-specialised (OES)

5. SAMPLING RESULTS

The sampling implementation process was carried out by a cross-organisational team of colleagues from the Australian Council for Educational Research, representatives from the Uzbekistan Ministry of Education, and the British Council Uzbekistan and UK.

From the defined comparison unit, 150 schools and 1,446 students were sampled for participation.

English Impact Uzbekistan achieved sampling levels within international standards. Findings can therefore be considered **representative of Grade 10 Uzbek students** studying English in publicly-funded schooling.

The target school population across the country was 100% of all publicly-funded schools throughout Uzbekistan. Schools for children with special needs were excluded (66 schools), along with 260 schools which had five or less students in the target grade. 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 5.2 below.

The overall rate of school level and within school exclusions (2.9%) was within the rate of 5% outlined in participation standard 1.3, therefore a high standard of participation was successfully achieved.

Table 5.1: Target population definition in Uzbekistan

Grade 10 equivalent in Uzbekistan	Years of formal Schooling	Average age at time of testing	Information about age of entry, promotion and retention
Grade 10	10	16.3	Children of 6-7 years old are admitted to Grade I of general education. As primary and general secondary education is compulsory in Uzbekistan, hiring relevant age groups in schools is carried out according to district or city municipalities based on a population census procedure. Students are retained only in exceptional circumstances and with parent consent. All other students are automatically promoted from one grade to the next.

Table 5.2: Coverage and exclusions

Coverage	Notes on coverage	School-level exclusions	Within-sample exclusions / refusals	Overall exclusions / refusals
100%	All schools in Uzbekistan public education system	0.3%	2.6%	2.9%

Table 5.3: 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
150	121	116	5	121

Table 5.4: 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
94%	1,446	0	37	1,409	78	1,331

As described in Chapter 3, a process of implicit stratification was used to ensure a proportional sample allocation across all implicit strata. The stratification variables applied to the Uzbekistan sample frame were administrative region, urban or rural location, and medium of instruction (English non-specialised (OES) and specialised (SES)).

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 150 schools were drawn, of which 121 were eligible for participation, as shown in Table 5.3. 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. Five substitute schools were used in Uzbekistan, due to non-response from five first sampled schools. There was no participation from two of the original sampled schools.

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 5.4 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 who took part in the assessment across the region was 1,331.

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 5.5 below. The table shows the actual number of schools and students in Uzbekistan 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 school frame. As shown, the population size estimated from the sample is smaller than the population size from the sampling frame. This reflects the fact that a significant number of schools had to be excluded, since 29 sampled schools out of 150 had to be excluded since no students from the target population were found there[30].

The English Impact Uzbekistan response rate is an important participation standard and indication of the successful implementation of the school and student sampling procedure. Table 5.6 shows the weighted school and student participation rates. The weight applied to each school corresponds to the number of schools and students that they represent in the entire population. Each student within each school had a weight equal to:

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

The weight applied can vary from school to school. The weighted response rates, which consider 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% of all sampled students across responding schools, with an overall participation rate after replacement of schools of 88%. In Uzbekistan this standard was reached even before the use of replacement schools. For sampled schools only the participation rate was 90.02%, with 93.95% after adding the replacement schools.

Table 5.5: Population and sample size

Uzbekistan population		Achieved sample population		
Schools	**Students**	**Schools**	**Students**	**Student population size estimated from sample**
9,075	445,927	116	1,331	283,464

Table 5.6: Weighted school and student participation rates

School participation		Student participation	Overall participation	
Before replacement	**After replacement**		**Before replacement**	**After Replacement**
96.01%	**100.00%**	**93.95%**	**90.20%**	**93.95%**

30 School tracking information indicated that in some instances when the test administrators reached the school, language programmes were not offered in English (only, e.g., German, French); this perhaps reflects limitations in the capacity of collecting up-to-date relevant data.

6. LANGUAGE LEARNING ENVIRONMENT

From the 1,446 students sampled for participation, 1,331 students participated in English Impact Uzbekistan; of these, valid questionnaire response data exists for 1,319 students. These participants completed a 53-item questionnaire translated into Uzbek. This included the motivation questions reported on separately in Chapter 6 of this report. Other questions provided background information on participating students, and information about their English learning experiences and environment, both in and out of school, with a particular focus on digital engagement. A brief demographic profile of the participating students is given, followed by a report on the proportions of responses given to questions about students' language learning environment.

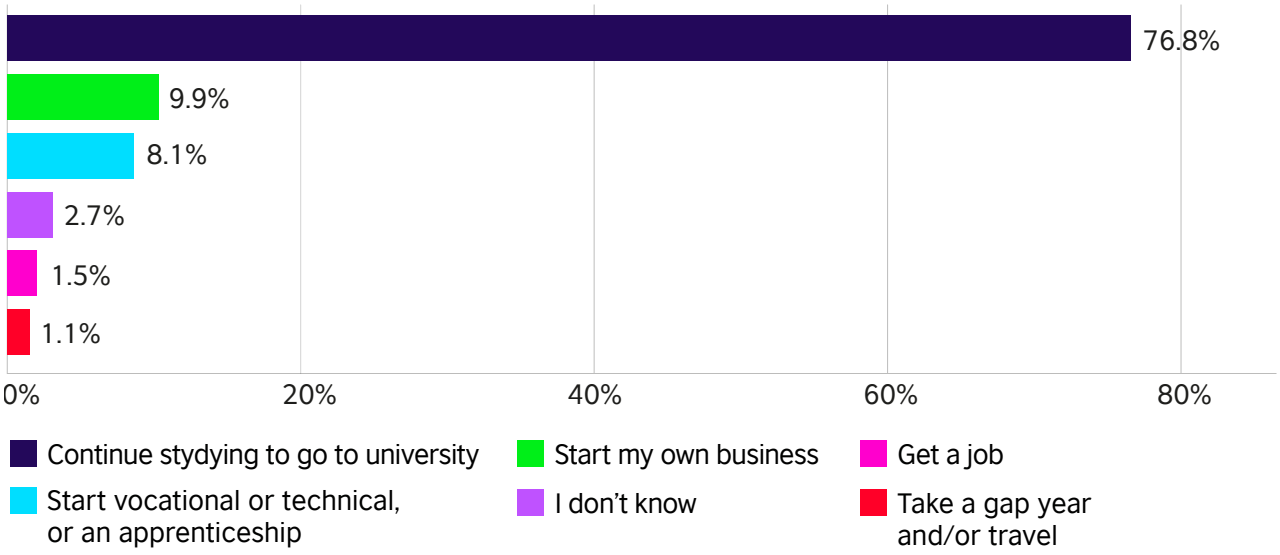
Table 6.1: Demographic variables of participating students from Uzbekistan

Demographic variables	Reported by participating students in Uzbekistan
Gender	56.0% Female; 44.0% Male
Age	Mean age of 16.3 years
Language most often spoken at home	83.2% Uzbek; 4.5% Russian; 12.1% other; 0.2% missing
Country of birth	99.0% Uzbekistan; 0.8% other; 0.2% missing
Attendance at pre-school	48.0% attended preschool; 52.0% didn't attend preschool

The gender ratio of female and male students was 56% and 44% respectively. As described in Chapter 3, the target average age of the target population was 15.5 years. The mean age reported by participating students was 16.3 years old, showing the accuracy of students sampled from the target population. A large majority of 83.2% indicated they spoke Uzbek most often at home, with Russian being a distant second at 4.5%. When asked to report their country of birth, over 99% selected Uzbekistan. The sample was a near even split between those who had attended pre-school (48%) and those who had not (52%).

The next question was intended to understand more fully the possible further pathways of school students from Uzbekistan in grade 11. All participating students were asked to report what they would like to do when they finished school. Over 75% suggested they wanted to go to university. Around 10% said they envisioned starting their own business while a slightly smaller number aspired to vocational or technical education. The remainder either did not know yet or mentioned getting a job or taking a gap year.

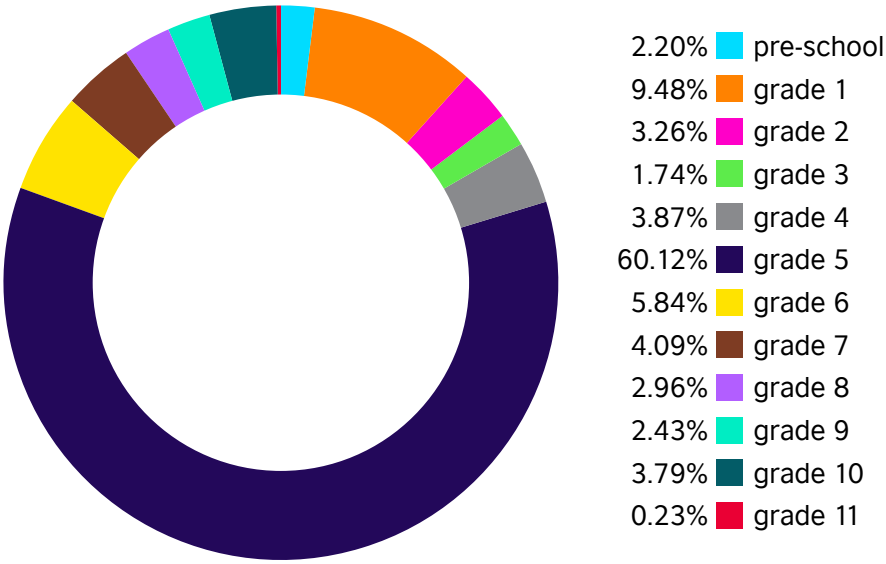
Figure 6.1: Intended future pathway of participants after compulsory schooling



Learning English at school

Only 2.2% reported they had started learning English as early as preschool. A majority of over 60% indicated they had begun learning in grade 5, with another 18.35% having started in primary school in a year before grade 5. Another 19.48% had started while in grades 7–9 while only 4.02% stated they had only started formal English lessons in grades 10 or 11.

Figure 6.2: The grade participating students started learning English in school



Students were asked to report if, in their current grade, they chose to study English: 71.72% said they chose to study English at school; 20.55% said it was a compulsory subject at school; and 7.73% said their parents suggested they study English.

When asked how long they spend each week at school learning English, the largest number, 36.6% said they spent between three and four hours per week studying English, as shown by Figure 6.3.

Figure 6.3: Whether participants reported to be studying English by choice or requirement

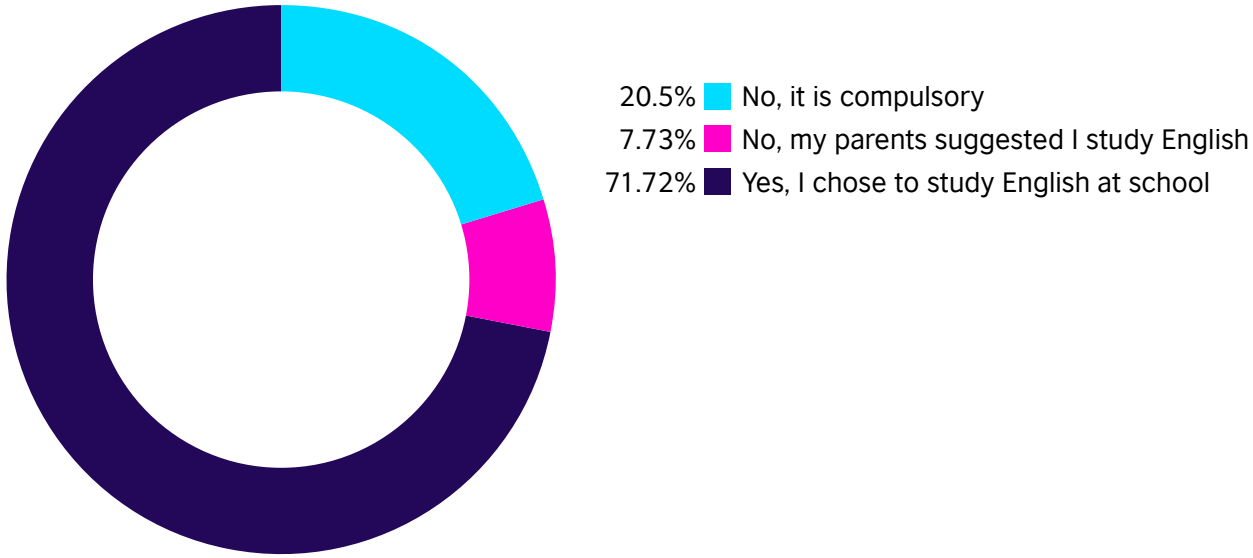
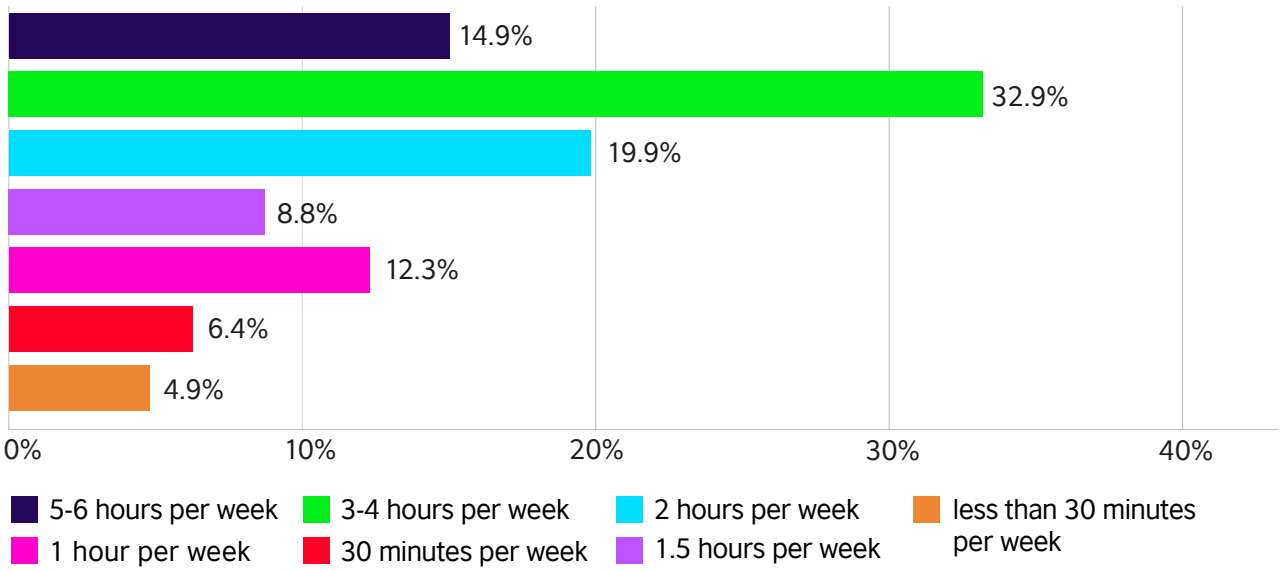
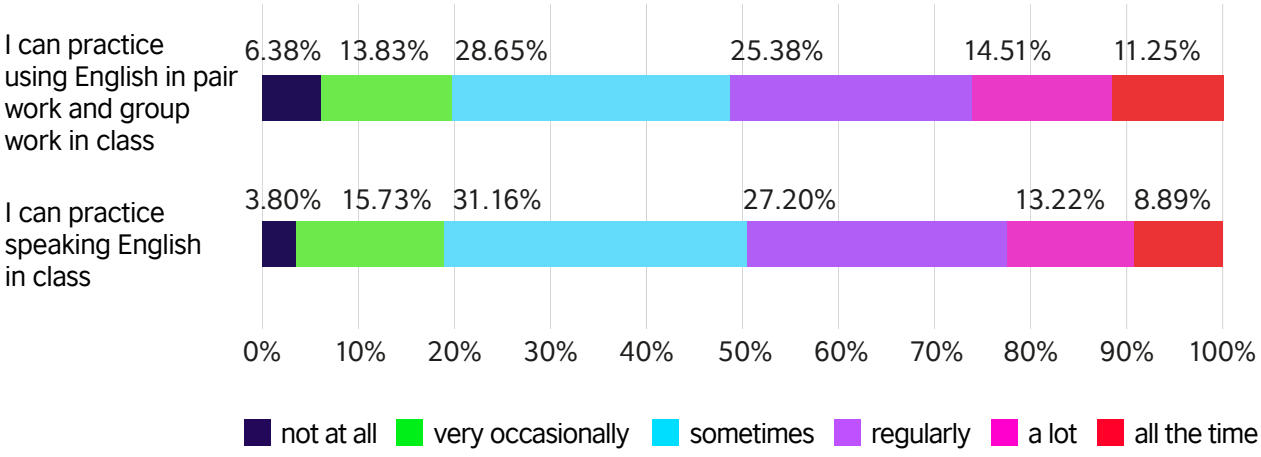


Figure 6.4: Time spent learning English, in dedicated English language lessons, in school per week



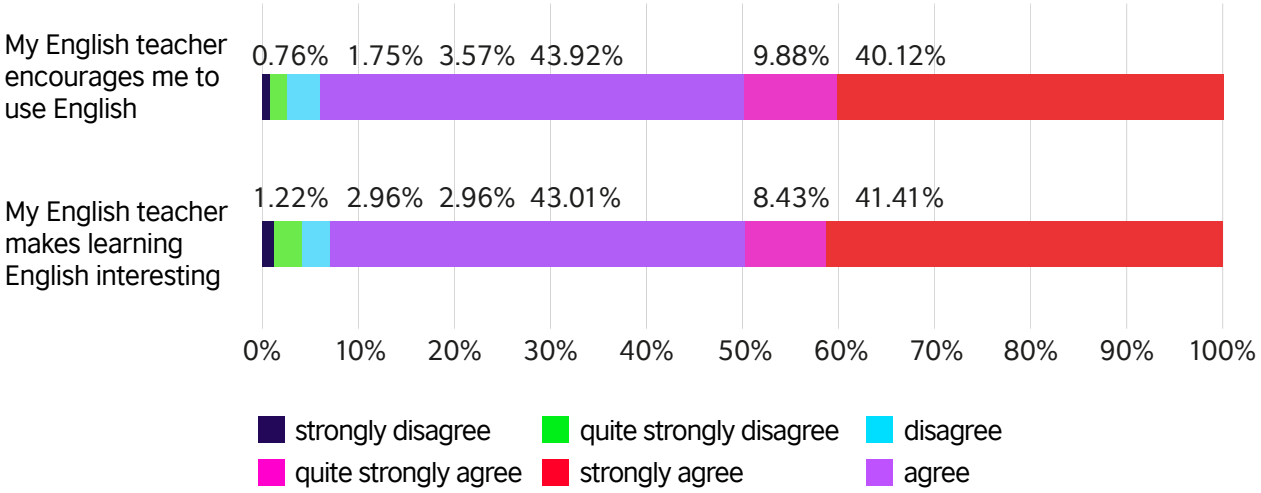
With respect to classroom experiences, it was of interest to see the wide variety of responses given to questions about the opportunity to speak English within the classroom, and also the use of pair work, as shown in Figure 6.5. Just over half the students reported being given the opportunity to practice English in pair work and group work in class regularly, a lot or all the time, and just under half the students reported to have similar chances to practice speaking English in class.

Figure 6.5: Pair work and opportunities to speak in class



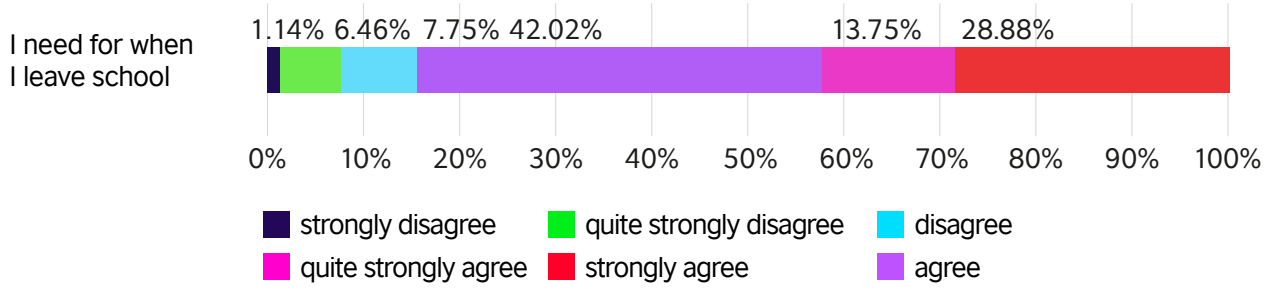
Participating students were also asked more detail about their perceptions of their English language learning experiences in the classroom, and its usefulness. Firstly, with respect to their experiences with their teachers, as shown in Figure 6.6, the overwhelming majority reported a positive experience in terms of teachers encouraging them to learn English and making English learning interesting.

Figure 6.6: Student perceptions of English teachers' approaches



It is clear that the majority of students were positive about the value of the skills they are learning, with more than 80% agreeing that they are learning the English skills needed for when they leave school, as shown in Figure 6.7. Some further insight into attitudes towards English language learning are given in the discussion of language learning motivations in Chapter 8.

Figure 6.7: Student perception that they are learning the English skills they need for when they leave school

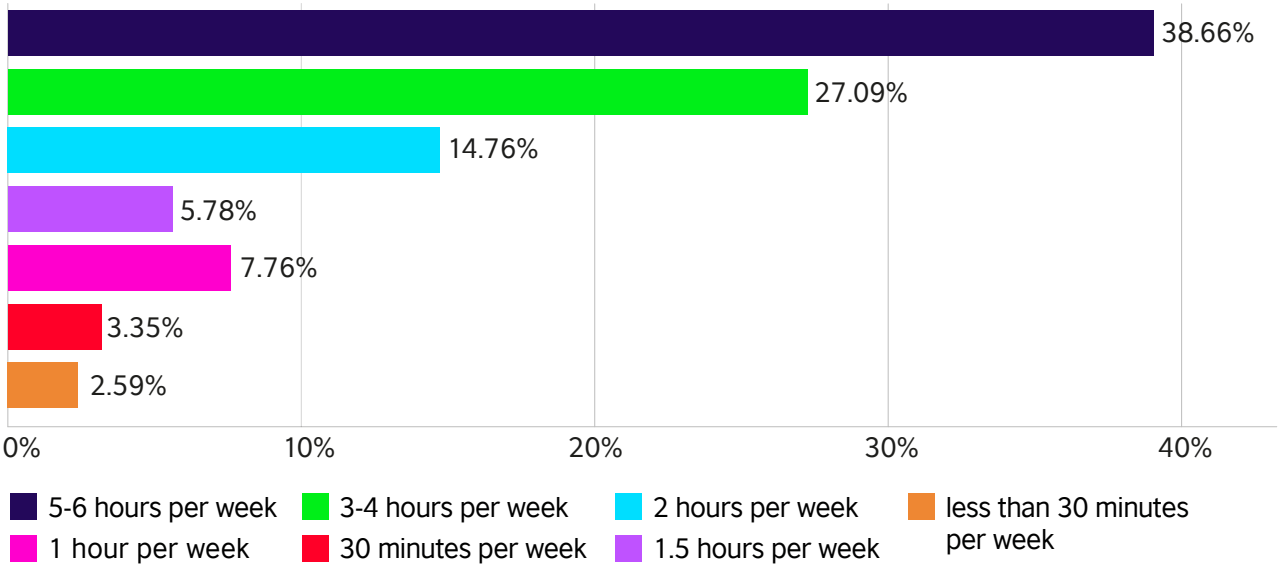


Learning English outside school

Participating students were also asked to provide information on studying English outside of school: 49.8%, said that they did choose to study English outside of school.

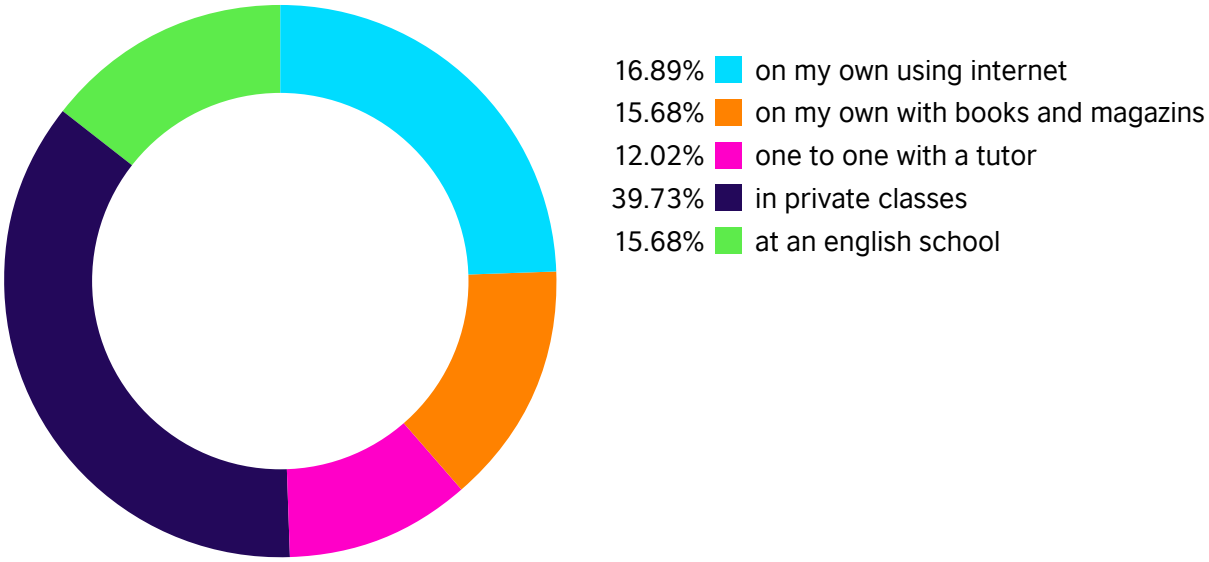
Those students who indicated they did choose to study English outside of school (657 in total) were asked to indicate how much time they spent learning English outside of school. Of this group, 38.3% spend between 5 and 6 hours a week on English language learning in addition to regular school hours. Most of the remaining respondents report spending between 2 and 4 hours per week on English after school.

Figure 6.8: Hours typically spent per week attending out-of-school-time lessons in English (subset of 49.8% of participating students who reported studying English outside of school)



The same segment of the participating student population was asked to indicate what activity best described their extracurricular language learning. Of these students, the largest proportion, 39.7%, said they studied English in private classes; 15.7% an English language school; 12% with a one-to-one tutor; 16.9% on their own using the internet; and 15.7% on their own using books and magazines.

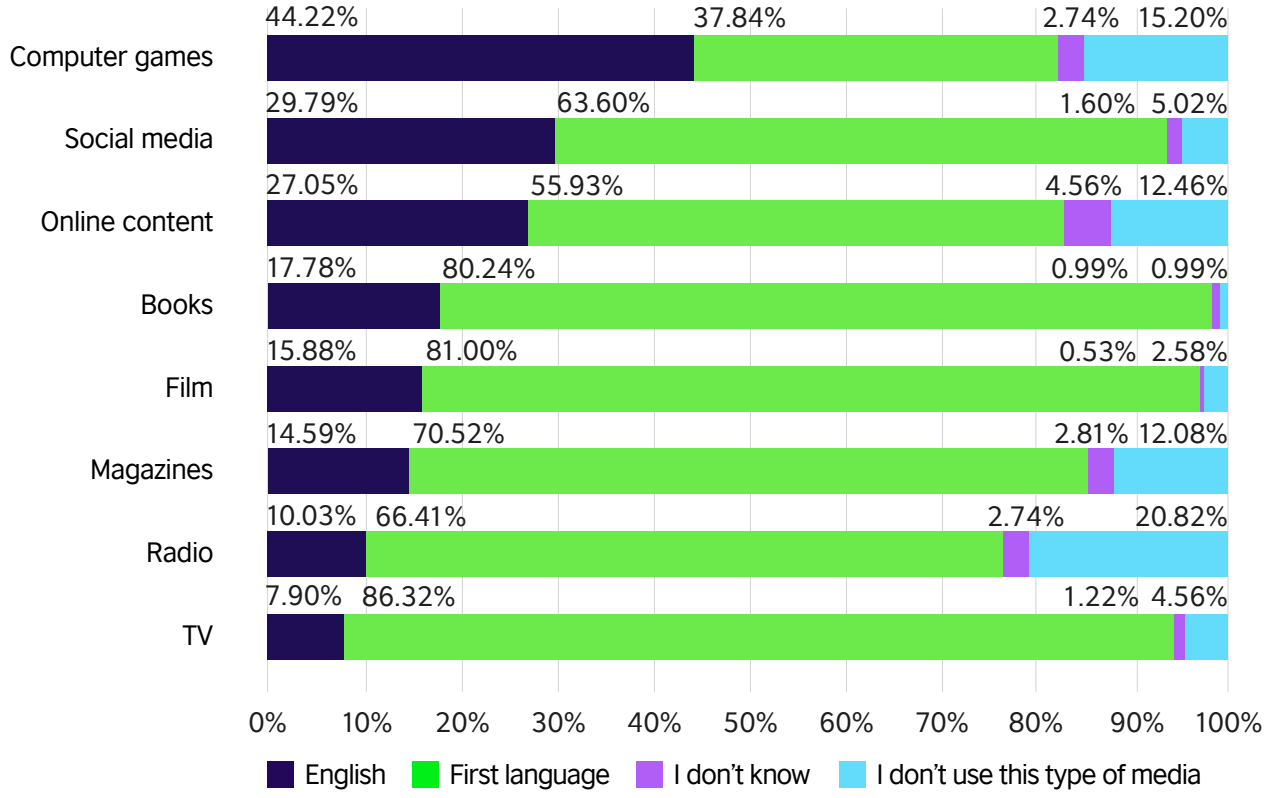
Figure 6.9: Means of studying English outside of school time (subset of 49.8% of participating students who reported to study English outside of school)



The use of technology and learning English

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 digitally native generation of learners that comprise the target population of English Impact Uzbekistan. To gauge and understand their language use when interacting with various types of media, they were asked whether they tend to consume different types of media most often in English or in their native language (note from Table 6.1 that for the vast majority, 83.2%, of participating students the native language is Uzbek, however this may also refer to Russian, or another language).

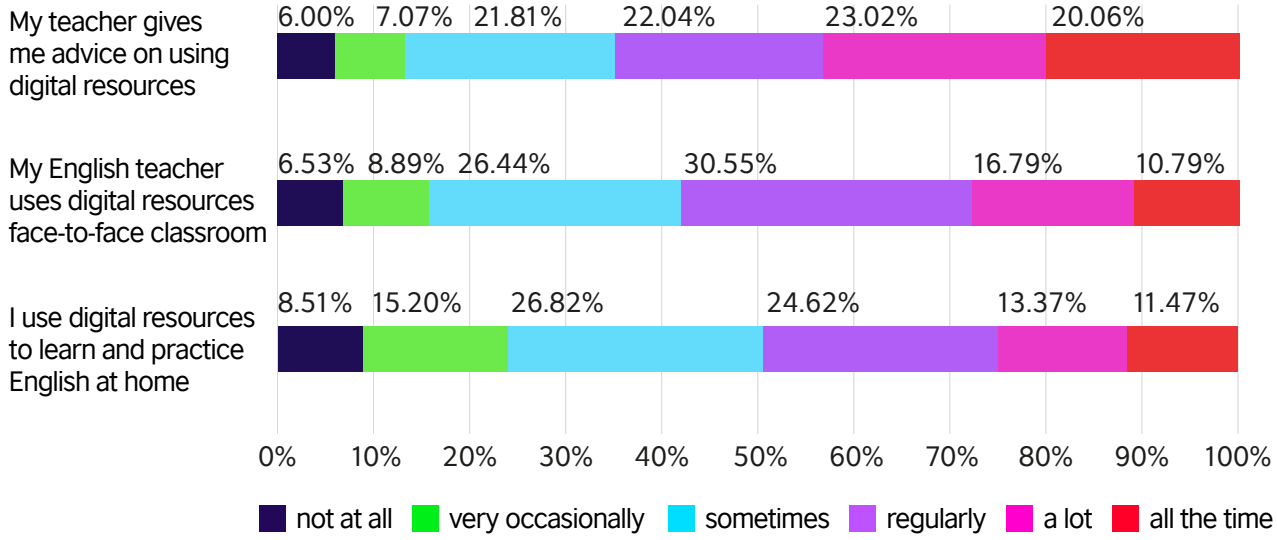
Figure 6.10: Language most regularly used to watch, read or listen to different types of media



The media most commonly consumed in English by the participants were computer games, with 44% of participants indicating that they regularly accessed these in English language medium. Other digital media regularly accessed in English were social media (almost 30% of participants) and other online content (around 27%). While Uzbek remained the most commonly-used language with each of these types of media, we can discern a pattern of English language use in technology for this demographic. Finally, with respect to digital engagement, there were three questions that specifically asked about use of, and engagement with, digital resources in their English learning practices, both at home and at school. The use of digital resources in English language learning is clearly widespread amongst the Uzbek students included in the sample, with only 6.5% of students reporting that their teacher never used digital resources in face-to-face classrooms, and only 8.5% stating that they did not use digital resources to practice English at home. This means that even at a conservative estimate, over 90% of students regularly use digital resources for English learning either in the classroom, or at home, or both. In addition, the vast majority, 94% of participating students, report that their teacher gives them advice on using digital resources.

Over 90% of students regularly use digital resources for English learning either in the classroom, or at home, or both.

Figure 6.11: Use of digital resources in English language learning



7. ASSESSMENT OUTCOMES

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. See Chapter 3 for further details.
Over half of the comparison unit population from Uzbekistan, 68.15%, achieved at A1 level in their overall English language test performance on the Common European Framework of Reference for Languages (CEFR). Of the participating population, 21.25% achieved at A2 level, 7.69% at B1 level, and 1.97% at B2 level as shown in Table 7.1 below.

Over half of sampled students, **68%** achieved an overall level of English of A1 on the Common European Framework of Reference for Languages (CEFR).

Table 7.1: Distribution of student population for overall test performance by CEFR level

CEFR Level	Overall English language test performance		
	Percentage (%)	**Standard error (%)**	**95% CI**
C	0.69	0.22	[0.3, 1.1]
B2	1.97	0.63	[0.7, 3.2]
B1	7.69	1.04	[5.7, 9.7]
A2	21.25	1.31	[18.7, 23.8]
A1	68.15	2.11	[64.0, 72.3]
A0	0.26	0.19	[0.0, 0.6]
Missing	0.00	0.00	[0.0, 0.0]
Total	**100**	-	

As described in Chapter 3, the 95% confidence interval is the region ± 1.96 standard error around the estimate and provides a measure of the certainty of the estimate. As shown in Table 7.2, the standard error of the mean estimates for all four skills are small, indicating a high level of precision for these estimates. Please note the individual test components are not calibrated to be directly comparable with respect to scale scores. CEFR level allocations shown in Tables 7.3 and 7.4 should be used as the basis for skills-level comparisons.

Table 7.2: Mean population score and CEFR levels by skills and overall achievement

	Grammar Vocab	Listening	Reading	Speaking	Writing	Overall
Mean	14.03	19.20	14.34	6.32	9.13	48.76
SE	0.281	0.342	0.326	0.473	0.527	1.553
95% CI	(13.48 - 14.58)	(18.53 - 19.87)	(13.70 - 14.98)	(5.40 - 7.25)	(8.10 - 10.16)	(45.72 - 51.80)

Receptive skills

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

The estimated distribution of the listening skill performance shows that 17.16% of participating students achieved at B1 level, while a high proportion, 67.28% achieved at A2 level for their listening skill.

Table 7.3: Estimated distribution of student population for listening and reading skills by CEFR level

	Listening component			Reading component		
CEFR level	**Percentage (%)**	**Standard error (%)**	**95% CI**	**Percentage (%)**	**Standard error (%)**	**95% CI**
C	0.54	0.19	[0.2, 0.9]	0.91	0.35	[0.2, 1.6]
B2	2.65	0.61	[1.5, 3.8]	1.52	0.49	[0.6, 2.5]
B1	17.16	1.16	[14.9, 19.4]	8.51	0.87	[6.8, 10.2]
A2	67.28	1.52	[64.3, 70.3]	52.98	1.28	[50.5, 55.5]
A1	11.48	1.15	[9.2, 13.7]	28.98	1.41	[26.2, 31.7]
A0	0.72	0.20	[0.3, 1.1]	6.87	0.77	[5.4, 8.4]
Missing	0.17	0.16	[0.0, 0.5]	0.24	0.18	[0.0, 0.6]
Total	**100**			**100**		

As shown in Table 7.3, 53% of students achieved at A2 level on the CEFR for English reading, and 29% at A1 level.

Productive skills

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

Participating students achieved their lowest performance scores for the productive skills. The distribution of speaking skills when referenced against the CEFR shows the largest proportion of students achieved at A0 level, 65.36% as shown in Table 7.4. Meanwhile, the estimated distribution of writing performance when referenced against the CEFR shows that 26.77% of all students achieved at A1 level. The estimated distribution shows a proportion of students, 21.05%, achieved above A1 level, at A2, B1, B2, or C level in writing skills.

Table 7.4: Estimated distribution of student population for speaking and writing skills by CEFR level

CEFR level	Listening component			Reading component		
	Percentage (%)	**Standard error (%)**	**95% CI**	**Percentage (%)**	**Standard error (%)**	**95% CI**
C	0.54	0.18	[0.2, 0.9]	0.54	0.20	[0.1, 0.9]
B2	1.51	0.38	[0.8, 2.2]	2.41	0.54	[1.3, 3.5]
B1	6.33	0.97	[4.4, 8.2]	7.22	0.97	[5.3, 9.1]
A2	8.37	0.90	[6.6, 10.1]	10.87	1.23	[8.5, 13.3]
A1	16.00	1.21	[13.6, 18.4]	26.77	1.60	[23.6, 29.9]
A0	65.36	1.96	[61.5, 69.2]	51.64	2.45	[46.8, 56.4]
Missing	1.90	0.57	[0.8, 3.0]	0.55	0.26	[0.0, 1.0]
Total	**100**			**100**		

Comparing achievement by gender

The ratio of participating female and male students was relatively evenly distributed at 55.7% female and 43.4% male.

As shown in Table 7.5, while female students consistently achieved higher mean scale scores for all test components, these differences were not statistically significant at the 95% level.

No significant difference in ENGLISH LANGUAGE test performance was found between female and male students.

Table 7.5: Mean performance score by gender, skills and overall achievement

Gender	Grammar and Vocabulary	Listening	Reading	Speaking	Writing	Overall
Female	14.33	19.35	14.59	6.45	9.33	49.55
Male	13.74	19.01	14.14	6.27	9.02	48.12
Difference (F-M)	0.59	0.34	0.45	0.18	0.31	1.43
Standard error	0.407	0.457	0.369	0.615	0.607	1.739
Comparison (95% confidence)	▶	▶	▶	▶	▶	▶

▲ Positive difference ▶ No difference ▼ Negative difference

Comparing achievement by urban and rural school location

One of the stratification variables selected to be applied to the Uzbekistan sample frame was an urban or rural marker for each of the participating schools. Analysis of the mean performance of schools classified as urban or rural highlights the different levels of performance at both types of schools.

Table 7.6: Mean performance score by school location, skills and overall achievement

School location	Grammar Vocab	Listening	Reading	Speaking	Writing	Overall
Urban	15.25	21.34	15.97	8.95	12.11	57.95
Rural	13.08	17.52	13.08	4.32	6.83	41.59
Difference (U-R)	2.17	3.82	2.89	4.62	5.28	16.36
Standard error	0.572	0.737	0.668	0.960	0.990	3.129
Comparison (95% confidence)	▲	▲	▲	▲	▲	▲

▲ Positive difference ▶ No difference ▼ Negative difference

As shown by in Table 7.6, the row describing comparative difference in the 95% confidence intervals, there was a positive and significant difference between urban and rural schools' performance in grammar and vocabulary, listening, reading, speaking, and writing skills, and overall, with urban schools outperforming rural schools in all cases.

Comparing achievement by region

The sample was divided by regions and independent samples were draw from each. Table 7.7 shows the mean overall scores per region and a comparison with the results of each of the other regions based on the significance of the differences between the main scores and their standard error.

Table 7.7: Pairwise overall mean score performance with 14 Uzbekistan provinces

TIER	Region	Overall mean Performance	SE	Navoiy region	Tashkent city	Samarkand region	Andijan region	Fergana region
1	Navoiy region	76.8	(8.59)		▶	▲	▲	▲
	Tashkent city	72.1	(4.42)	▶		▲	▲	▲
2	Samarkand region	53.1	(5.08)	▼	▼		▶	▶
	Andijan region	50.4	(6.64)	▼	▼	▶		▶
	Fergana region	48.5	(4.17)	▼	▼	▶	▶	
	Namangan region	47.1	(5.47)	▼	▼	▶	▶	▶
	Tashkent region	46.7	(5.32)	▼	▼	▶	▶	▶
	Korakalpak AR	45.4	(6.32)	▼	▼	▶	▶	▶
	Bukhoro region	44.0	(3.34)	▼	▼	▶	▶	▶
	Khorazm region	42.2	(8.35)	▼	▼	▶	▶	▶
3	Jizzakh region	40.1	(4.15)	▼	▼	▼	▶	▶
	Sirdaryo region	35.3	(5.78)	▼	▼	▼	▶	▶
	Kashkadaryo region	38.5	(2.03)	▼	▼	▼	▶	▼
	Surkhandaryo region	34.4	(3.67)	▼	▼	▼	▼	▼

▲ Province on the left higher than province at the top
▼ Province on the left lower than province at the top
▶ No difference between the two provinces

Based on the pairwise comparison of all the regions, they can be divided into three distinct tiers. Inside each tier, there is no statistically significant difference between the scores of each of the regions that compose it. These tiers are:

- Tier 1 which consists of the Navoiy Region and Tashkent city: Both have a performance that is significantly higher than the rest of the regions and well above the average overall performance of the country (48.76, as shown in Table 7.2). Based on the result, Navoiy region has the higher regional score of all the country, but given the SE of the estimates, the difference with the mean performance of the second highest region, Tashkent city, is not statistically significant.
- Tier 2 which consists of the regions of Samarkand, Andijan, Fergana, Namangan, Tashkent, Korakalpak, Bukhoro and Khorazm: These regions have the characteristic that their performance is significantly below that of the regions in Tier 1. However, their performance is within the average of the country performance
- Tier 3 which is formed by the Jizzakh and Sirdayo, Kashkadaryo, and Surkhandaryo regions: These regions performed significantly lower than regions in Tier 1 and some regions from Tier 2. These regions perform significantly lower than the national average.

Comparing achievement by English specialisation

The final stratification variable able to the Uzbekistan sample frame was English specialisation. The mean scale score of English specialised schools (SES) was 22.25, while those for non-specialised schools was 59.01. The difference was statistically significant, as well across all skill levels. However, it must be noted that of all the participating schools, only one school belongs to the category of English specialised schools (SES), given the small proportion of schools in this category present in the whole population.

Table 7.8: Mean performance score by English specialisation, skills and overall achievement

School Type	Grammar Vocab	Listening	Reading	Speaking	Writing	Overall
English non-specialised schools	14.06	19.28	14.40	6.38	9.20	49.01
English specialised schools	11.00	10.83	8.83	0.25	2.33	22.25
Difference (OES-SES)	3.06	8.44	5.56	6.13	6.86	26.76
Standard error	0.280	0.335	0.319	0.471	0.522	1.529
Comparison (95% confidence)	▲	▲	▲	▲	▲	▲

▲ Positive difference ▶ No difference ▼ Negative difference

8. ENGLISH LANGUAGE LEARNING MOTIVATION

The analysis presented in this chapter explores two key areas:

- the levels and character of language learning motivation for different groups of students (split by gender)
- relationships between each of the motivational scales and English language proficiency.

In total, 1,319 participants were included in the analysis. Of these, 1,296 fully completed both the questionnaire and the Aptis test. Two students finished the test but had some questionnaire data missing. Similarly, a further 21 completed the questionnaire but did not finish all parts of the test.

The student context questionnaire records a good range of responses, which indicates that participants took the survey seriously and gave it their attention and consideration. The responses in the motivational part were, however, skewed towards the positive options, with 'agree' and 'strongly agree' being the most frequently chosen. This is important, since the analysis described in the upcoming chapter accounts for nuances in the patterns of response to this part of the survey.

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 results of an initial exploratory factor analysis showed that the international orientation did not emerge as a separate factor. This indicates that responses to this set of questions did not reflect a common underlying trait amongst the participating students.

It is not clear exactly why this is, but it could reflect that the students do not have a consistently expressed desire to connect with other English users online, or it may be that the set of questions was misinterpreted by the students resulting in a disconnect in their response patterns. This latter issue is a risk of using translated questionnaire items. Additionally, two further items were removed from analysis: INSTR4 and EX2. This indicates that these questions did not function as intended. Please cross-reference terms with Table 3.2 above.

The measurement model carried forward for confirmatory analysis has the structure shown in blue in Figure 8.1 below. This initial model showed this structure to have a good fit to the data according to accepted thresholds for CFA models (CFI = 0.934; TLI = 0.923; RMSEA = 0.046). It was found that the model could be improved slightly by allowing some of the error terms (i.e. residual error) for individual questions to correlate with each other. These relationships are indicated in green on Figure 8.1. 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[31]. The final measurement model therefore included the seven correlated error variances listed in Table D.1 in Appendix D. The model provides a good fit for the data (CFI = 0.952; TLI = 0.942; RMSEA = 0.040). The full output from the final model can be found in Appendix C.

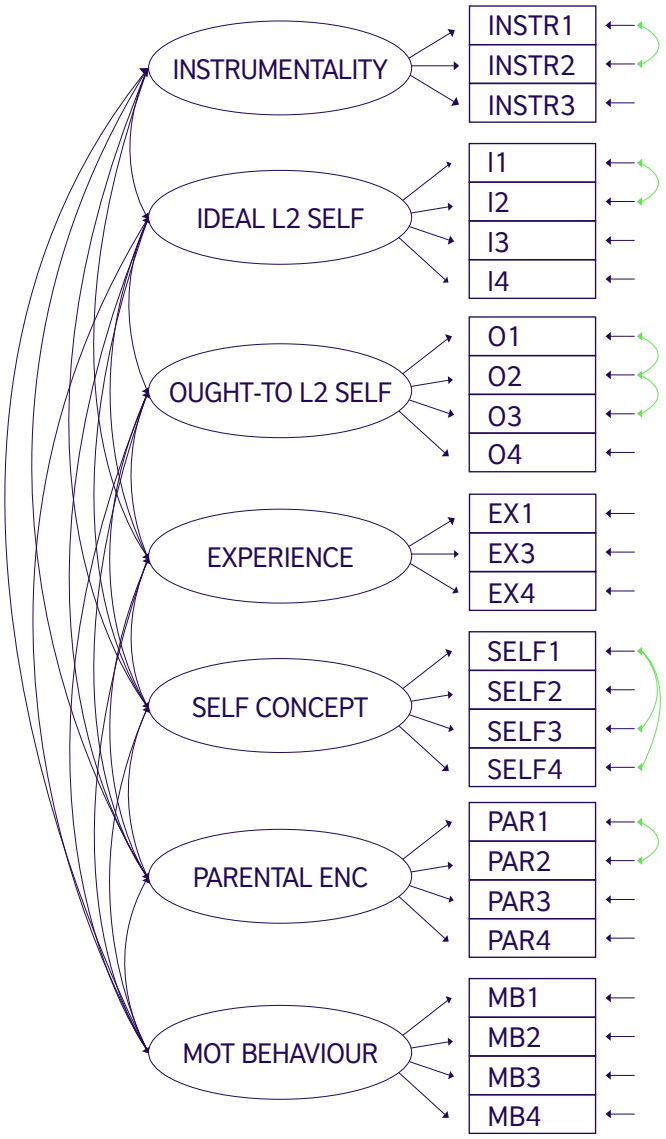
31 See Brown (2015, pp. 157–162) for a fuller discussion of correlated measurement error and the implications in CFA.

Figure 8.1: Measurement model

Double-headed arrows represent covariances between each of the latent variables:

Boxes show the observed variables (question responses)

Green double-headed arrows indicate correlated error terms



Measurement model

Details of the error covariances included in the model are given in Appendix D.

Structural model

Having established the viability of the measurement model, it is insightful to examine the association between the latent variables (representing each of the motivational traits) as estimated by the model. Standardised covariances (correlations) estimated by the model are given in Table 8.1. There was some variation in the values but all of them could be considered strong (Cohen, 1988).

Two latent variables are slightly 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 very clearly focus on external pressures connected with studying English. The first one refers to levels of parental views and expectations related to learning English, and the second one relates to pressures from broader environmental 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 study English (Ryan & 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 & Kormos, 2009; Iwaniec & Ullakonoja, 2016; Taguchi, Magid & Papi, 2009). This is in contrast with more internalised motivational traits, 'ideal L2 self', 'self-concept', and the traits reflecting 'instrumentality' and 'language learning experience' which all have much stronger relationships with motivated learning behaviour.

Table 8.1: Standardised covariances between latent variables in the final model

Motivational scales (short ref)		Estimate
IDEAL	EXPER	0.829**
INSTR	EXPER	0.873**
INSTR	IDEAL	0.880**
MOTIV	IDEAL	0.821**
MOTIV	INSTR	0.839**
MOTIV	EXPER	0.840**
OUGHT	IDEAL	0.793**
OUGHT	EXPER	0.749**
OUGHT	MOTIV	0.768**
OUGHT	INSTR	0.915**
PAREN	EXPER	0.732**
PAREN	MOTIV	0.731**
PAREN	IDEAL	0.791**
PAREN	OUGHT	0.879**
PAREN	INSTR	0.915**
SELF	OUGHT	0.694**
SELF	PAREN	0.622**
SELF	MOTIV	0.839**
SELF	INSTR	0.701**
SELF	EXPER	0.741**
SELF	IDEAL	0.724**

** p<.01 *p < .05

Multi-group analysis findings

Comparative analyses were conducted to compare female and male learners in the data.

The final sample included 1,319 students and was slightly skewed towards females as it contained 738 female and 580 male respondents. One participant did not identify their gender.

Female and male student populations' motivation

It is prudent not to assume identical motivations drive female and male learners 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; Iwaniec, 2019; Jimenez Catalan, 2010). Considering that motivation is shown to affect language learning achievement (Hsieh & Kang, 2010; Kim & Kim, 2014; Marsh & 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 gender groups were found between the levels of motivation reported on four of the seven motivational scales included in the analysis. See Appendix D for full statistical output. In all cases, female students reported higher levels of motivation than their male peers (significance level from t-test indicated in parenthesis):

- motivated behaviour ($p=.001$)
- English self-concept ($p<.001$)
- ideal L2 self ($p=.026$)
- language learning experience ($p=.002$).

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

- invest more effort in language learning
- have more robust visions of themselves as successful English users in the future
- have more positive language learning experiences, conceptualised as language learning attitudes
- perceive themselves as better language learners.

These results are consistent with findings from previous empirical studies (see Iwaniec [2015, 2019] for reviews). 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 effects for the perceptions of masculinity of males learning a foreign language (Carr & Pauwels, 2005b).
- Language proficiency is perceived as more directly relevant for future careers of female than male learners (Clark, 1998); the perception that is even more augmented by the perceived limited choice of professional careers for females (Norton & Pavlenko, 2004).
- Females are perceived as having a stronger preference for, and value communication more, than male learners (Chavez, 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).

No statistically significant gender differences were found for three motivational scales: instrumentality, ought-to L2 self and parental encouragement These suggests that both genders perceive English as important for their professional career and feel similar levels of pressure and encouragement from their environment, including their parents, to study English.

Relationships between motivational scales

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

- parental encouragement with ideal L2 self (5.940 on 1 d.f., p=.025)
- instrumentality with parental encouragement (5.026 on 1 d.f., p=.025)
- instrumentality with ideal L2 self (6.919 on 1 d.f., p=.01)
- ought-to L2 self with instrumentality (4.268 on 1 d.f., p<.05).

This results in a model with good overall fit statistics (CFI= 0.944; TLI = 0.940; RMSEA = 0.041). A short description of the analysis, and adjusted chi-square calculations are given in Appendix D, and a summary of the covariance estimates for this gender-specific model are given in Table 8.2.

Table 8.2: Latent covariances allowed to vary between gender groups (standardised estimates)

Covariance (using short ref)	Estimate
PAREN WITH IDEAL	
Female model	0.753**
Male model	0.844**
INSTR WITH PAREN	
Female model	0.814**
Male model	0.894**
INSTR WITH IDEAL	
Female model	0.877**
Male model	0.890**
OUGHT WITH INSTR	
Female model	0.949**
Male model	0.883**

** p<.01 *p < .05

As can be seen from the estimates column in Table 8.2, in three out of four 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
- parents of male learners are more likely to play a role in encouraging them to learn English to enhance their future employment opportunities than parents of 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.

In one case, the covariance is stronger for female learners than male ones. It appears that the female learners' vision of what they are expected to do with their English skills is more strongly related to using English to enhance their job opportunities than for male learners.

Links between motivation and proficiency

In this section, the relationship between Uzbek 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 'a 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 the 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 Uzbek students was measured at a time they took the proficiency test. However, motivation is dynamic and changes over time. Hence, while 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 therefore, 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 the Aptis test scores and motivational scales in general and divided by gender. 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 (fscores) derived from the CFA described above. As only two students achieved A0 level of proficiency, this group was merged with those who achieved A1 level. Similarly, due to small numbers of students achieving levels B2 (24 students) and C (8 students), these two groups were merged with students who achieved B1 (99 students). Overall, the students were divided into three groups A1- (A0 and A1) with 900 students, A2 with 288 students, and B1+ (B1, B2, C) with 131 students. Figure 8.2 shows the relative patterns of mean factor scores for students falling in each group, while Figure 8.3 summarises where significant differences between these groups were found. As can be seen from both figures, learners in the B1+ group have the highest scores on all scales. Group A2 has higher scores on all the scales than group A1-. However, further analysis revealed that only differences between groups A1- and B1+, and A1- and A2 are significant. This suggests the existence of two broad groups of learners, the more motivated learners who have A2 proficiency or higher, and the less motivated group whose members have lower proficiency.

Figure 8.2: Motivational profiles for learners at different levels of proficiency

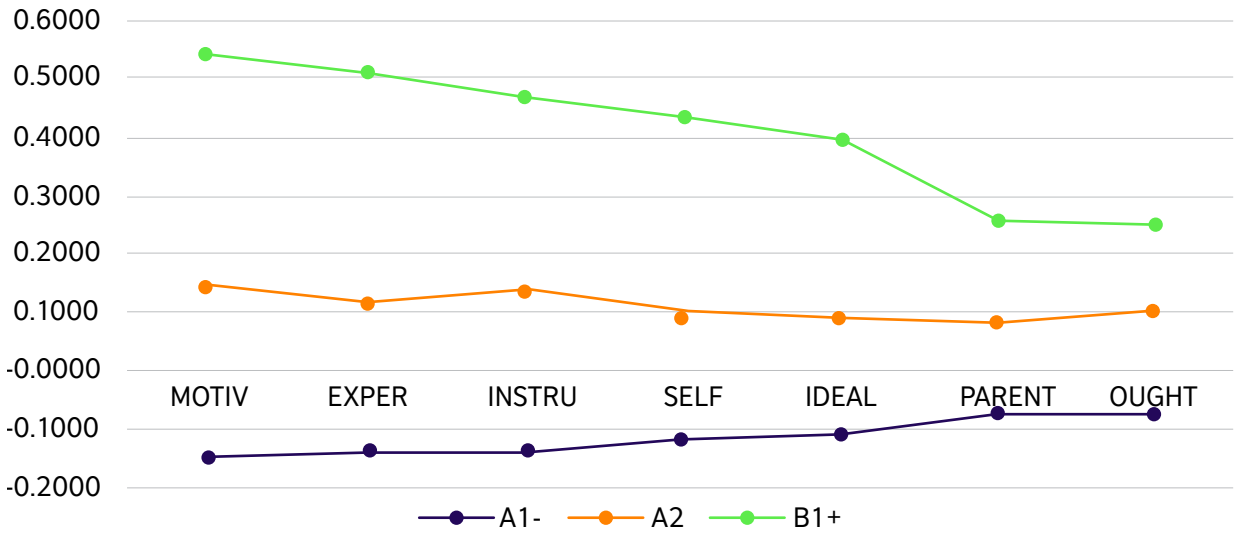


Figure 8.3: Visual representation of significant differences on motivational scales for learners at different levels of proficiency

CEFR level	English self-concept	Ideal L2 self	Language learning experience	Instrumentality	Motivated behaviour	Parental encouragement	Ought-to L2 self
B1+							
A2							
A1-							

Key: Darker colour = higher factor score
Different shades = statistically significant difference

On the whole, the motivational scores 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. More proficient students also report greater levels of parental encouragement than their less proficient peers.

Correlations between proficiency and motivation

The results presented in Table 8.3 show the strength of correlations between the overall scale score and the factor scores for motivational scales. The correlations are listed in descending strength.

Table 8.3: Correlations between motivational variables and overall scale score

Motivational scale	Total score
Language learning experience	.285**
Ideal L2 self	.276**
English self-concept	.256**
Motivated behaviour	.255**
Instrumentality	.225**
Parental encouragement	.166**
Ought-to L2 self	.140**

** p<.01 *p < .05

All motivational variables are positively related to proficiency, although all the correlations could be classed as weak. Nevertheless, it is clear that there are differences in the strength of correlation between different aspects of motivation and proficiency, with more internalised aspects of motivation, such as language learning experience, ideal L2 self, English self-concept and motivated behaviour being more closely related to proficiency, as compared to the more externalised aspects of motivation, such as parental encouragement and the ought-to L2 self.

Table 8.4 meanwhile shows correlations between motivational variables and different components of the proficiency test. Compared to the overall proficiency scores, correlations with scores on individual components tend to be 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, writing, and grammar and vocabulary, tend to be slightly higher than with reading, whereas those with listening tend to be the lowest. These differences are, however, small. Overall, for the individual aspects of proficiency, the order of strength of correlations with different motivational patterns remains largely the same. This means that language learning experience, ideal L2 self and English self-tend to be more strongly correlated to all the aspects of proficiency than the more externalised aspects of motivation, i.e. parental encouragement and the ought-to L2 self.

Table 8.4: Correlations between motivational variables and scale scores for individual Aptis components

Motivational scale	Listening	Reading	Speaking	Writing	Grammar & vocabulary
Language learning experience	.188**	.232**	.268**	.289**	.253**
Ideal L2 self	.195**	.217**	.250**	.286**	.237**
English self-concept	.175**	.192**	.243**	.266**	.222**
Motivated behaviour	.165**	.204**	.244**	.262**	.227**
Instrumentality	.143**	.180**	.212**	.235**	.201**
Parental encouragement	.097**	.139**	.155**	.177**	.161**
Ought-to L2 self	.075**	.108**	.139**	.154**	.130**

** p<.01 *p < .05

Below is the discussion of findings scale by scale.

- Language learning experience: Positive language learning experience, measured by attitudes to language learning, is most strongly related to proficiency in Uzbekistan. A possible explanation for this might be that interesting and rewarding learning experiences can help sustain effort investment in the short run (Ford, 1992) and create a positive outlook for future activities. Hence, it is vital that English classes are a source of such positive experiences.

- Ideal L2 self: The ideal L2 self is also a key to heightened proficiency scores. This reflects the personal language goals of the students and 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 the learners have time to reflect on what they will be using English for in the future. This finding is actually in contrast to the finding reported by Moskovsky, Racheva, Assulaimani and Harkins (2016). However, as numerous studies reported a link between effort investment and the ideal L2 self (Iwaniec, 2014; Kormos, Kiddle & Csizér, 2011; Taguchi et al., 2009)2005, and more recently Dunn and Iwaniec (2021) reported a link between ideal L2 self and proficiency, it is unsurprising.

- English self-concept: This trait, which reflects students' self-confidence in learning English is also related to proficiency scores. This means that Uzbekistani students appear to make broadly correct evaluations of their own ability to learn and speak English. This finding is unsurprising as previous studies (Dunn & Iwaniec, 2022; Hsieh & Kang, 2010; Hsieh & Schallert, 2008; Mills, Pajares & Herron, 2007) reported a link between self-concept and achievement. As well as reflecting performance, it is also understood that ensuring that students' English self-concept is positive can be a first step towards higher proficiency in a circular cause and effect relationship (Bandura, 1997).

- Motivated learning behaviour: This trait is also related to proficiency. This means that there is a general link between the reported level of effort invested in language learning and students' 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 only might not be sufficient to fully explain the correlation between effort investment and motivation.

- Instrumentality: Having clear language learning goals, such as using English in a future career (instrumentality) is weakly 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 a form of short meetings with local people discussing how English is useful in their jobs (for example, small business owners, builders, those who work or have worked abroad).

- Parental encouragement: The correlation between parental encouragement and proficiency is positive but low. This is unsurprising, considering that the students from the current study are 15-year-olds, which is a stage when teenagers value newly gained independence from their parents. 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 (Dunn & Iwaniec, 2021; Iwaniec & Ullakonoja, 2016; Iwaniec, 2015; Kyriacou & 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 rather weak. This means that the social expectations and pressure from the external environment is not a force motivating students to learn English. This finding is unsurprising as previous studies have also pointed to a limited role of the ought-to L2 self in many contexts (Csizér & Kormos, 2008a, 2008b, 2009, Dunn & Iwaniec, 2021).

Motivation-proficiency correlations by female and male populations

The results of correlational analysis of the cohort split along gender lines (Table 8.5) show that even though the correlations between motivational variables and overall proficiency appear, on the whole, stronger for females, there are no significant differences between the two groups. Interestingly, the relationship between proficiency and both parental encouragement and ideal L2 self is stronger for male than female learners, albeit not significantly.

Table 8.5: Correlations between overall proficiency and motivational variables, by gender

Motivational scale	Female	Male	P-value[†]
Language learning experience	.296**	.252**	.392
Ideal L2 self	.270**	.276**	.907
English self-concept	.283**	.229**	.299
Motivated behaviour	.274**	.227**	.902
Instrumentality	.231**	.207**	.650
Parental encouragement	.142**	.192**	.355
Ought-to L2 self	.176**	.092*	.124

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

The patterns observed when discussing the links with overall proficiency are mirrored when correlations with individual components of the test are explored (Table 8.6). In more detail, it can be seen that while the differences between two genders remain small, in some cases, the correlation did not reach significance, namely in the case of parental encouragement and listening for girls, and the ought-to L2 self and listening/reading for boys. This is not surprising, however, considering that these two are the most externalised aspects of motivation.

Table 8.6: Correlations between motivational variables for individual Aptis components, by gender

Motivational scale	Listening		Reading		Speaking		Writing		Grammar & vocabulary	
	female	male	female	male	female	male	female	male	female	male
Language learning experience	.195**	.156**	.241**	.202**	.265**	.255**	.306**	.260**	.253**	.237**
Ideal L2 self	.178**	.211**	.206**	.223*	.244*	.249**	.286**	.283**	.210**	.264**
English self-concept	.204**	.139**	.209**	.176**	.255**	.235**	.293**	.240**	.234**	.207**
Motivated behaviour	.186**	.129**	.222*	.181**	.244**	.243**	.281**	.234**	.235**	.211**
Instrumentality	.142**	.138**	.186**	.163**	.211**	.199**	.242**	.220**	.193**	.207**
Parental encouragement	.067	.133**	.124**	.157**	.133**	.179**	.152**	.206**	.125**	.208**
Ought-to L2 self	.100**	.038	.140**	.065	.165**	.109**	.188**	.107**	.145**	.108**

9. OVERVIEW OF FINDINGS

The English Impact project gives a snapshot of English language capability of grade 10 Uzbek students learning English in a formal setting. All participating students were studying at least 90 minutes of English per week within the publicly funded education system during project implementation at the end of 2021. Employing a two-stage cluster sample, a nationally representative range of schools and students participated in the study, with overall participation indicating 100% of school numbers achieved and 94% of students. This means that the information given by the study is highly relevant for informing policy decisions, and to act as a robust baseline information against which future comparisons can be drawn.

The survey findings give detailed information about the levels of English language proficiency across four skills (listening, reading, speaking, and writing) plus grammar and vocabulary using a fully validated English language test developed especially for teenage learners of English (Aptis for Teens). Accompanying this are findings which indicate student attitudes towards their language learning environment and digital engagement, plus a full analysis of the levels and relationships between differing aspects of student motivation, a series of interlinked traits which are understood by second language acquisition researchers to influence language learning.

With respect to English language proficiency, 68% of sampled students achieved an overall level of English of A1 on the Common European Framework of Reference for Languages (CEFR), and 21% a level of A2. These levels are both classed as a 'basic user' of English language, according to the CEFR 2001 global scale (Council of Europe, 2001; and see Appendix A). The majority of learners are at level A1, the first rung on the CEFR 2001; the Council of Europe emphasises that this 'is not the lowest imaginable level of proficiency in an additional language' but rather 'the point at which the learner can interact in a simple way, ask and answer simple questions about themselves, where they live, people they know, and things they have, initiate and respond to simple statements in areas of immediate need or on very familiar topics' (Council of Europe, 2020). In terms of the individual language domains, the receptive language skills of listening and reading were the strongest areas for the students, with 67% and 53% achieving an A2 in these test components respectively, and 99% and 93% achieving A1 or above. Productive skills of speaking and writing were comparatively lower, with 33% and 48% achieving A1 or above in these skill areas.

In comparing groups, no significant difference in performance was found between female and male students, urban schools were shown to outperform rural schools across all language domains and overall. For the regional comparisons, three tiers of performance were identified, with the Navoiy Region and Tashkent city as the strongest performers in Tier 1. It was not possible to conduct a meaningful comparison between English specialised schools (SES), and non-specialised schools, since there was only a single specialised school selected for inclusion in the school sample. If a more formal statistical comparison were to be made between specialised and non-specialised schools in future studies, this would require an oversampling of schools that meet this criterion.

An important aspect of the English Impact project is that it does not simply take into account current proficiency levels, but rather integrates this information with insights into attitudes of the students towards their language learning as reflected by the survey information. Marrying together this information is to take a broader perspective on the outlook of the students and the potential for developing English language proficiency. It also provides empirical insights based on second language acquisition theory upon which to focus future teaching approaches as well as policy-level decisions.

In terms of language learning motivation, the findings show relationships with proficiency, with language learning experience – operationalised as positive language learning

attitudes – found to be the trait most closely related to achievement. Other more internalised aspects of motivation, such as ideal L2 self, English self-concept and motivated behaviour also appear to be more strongly related to proficiency than the more externalised aspects, including learning English for utilitarian purposes, such as enhanced job prospects, parental encouragement, or the sense of obligation (ought-to L2 self). Although relationships between language learning motivation and proficiency is only a small part of the full picture of the language learner, a pattern of the relationship between English language proficiency and motivation levels emerges from the study. Overall, however, the correlations with proficiency and all aspects of motivation are relatively weak. This means that there may be some other variables that have not been accounted for in this investigation that prevent 15-year-olds in Uzbekistan to achieve their potential. Further investigation would be necessary to establish what these intervening variables might be.

Gender-wise, boys reported lower levels of motivation than girls on four scales. Importantly, these differences are on the more internalised aspects to motivation that are more closely related to proficiency, rather than on the externalised aspects. This suggests that in crucial aspects, boys are less motivated than girls to learn English. It also appears that parents play a stronger role in the motivation of male than female learners. These types of insights might be useful when planning to encourage development of English language skills.

An interesting finding to arise from the study is that one of the hypothesised motivational traits – international orientation – did not emerge as a clearly defined factor among the students in the study. This refers to the goal of learning English for the purpose of communicating with other users of English, either face-to-face or online (Yashima, 2000; Iwaniec, 2014). Along with instrumentality, this variable has been shown to be a more valuable external trait of motivation in explaining proficiency outcomes in other research. It would therefore be interesting to further explore why this scale did not function as expected for the students, and how international orientation affects English learners in Uzbekistan. A focus on developing international outlook and engagement could prove beneficial in the enhancement English language learning (cf. Botes et al., 2020).

Further important findings relate to the Uzbek students' engagement with digital media. Nearly 30% of participants reported using social media and other online content in English language medium. There was widespread use of digital resources to aid English language, with over 90% of students regularly using digital resources for English learning either in the classroom, or at home, or both.

10. IN CONCLUSION

The aim of this research was not to create a single score to show the levels achieved for English language teaching and learning in Uzbekistan. The detailed description of the research processes and outcomes presented in the English Impact Uzbekistan 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.

A number of factors came together to produce this ground-breaking research. The research design, combining ACER's sampling expertise and the British Council's knowledge and experience in English language assessment. Alongside the concerted effort made by local British Council teams working with the ministry of public education in Uzbekistan and the regional administrative centres to collect all required data for sampling, this 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 a constructive relationship with teachers, schools and students participating in the research process was 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

Bachman, LF (1990) Fundamental considerations in language testing. Oxford, England: Oxford University Press.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W. H. Freeman & Company.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? Educational Psychology Review, 15(1), 1–40.

Botes, E., Gottschling, J., Stadler, M., & Greiff, S. (2020). A systematic narrative review of International Posture: What is known and what still needs to be uncovered, System, Vol 90. doi.org/10.1016/j.system.2020.102232

Byrne, B. M. (2012). Structural Equation Modelling with Mplus: Basic concepts, applications, and programming. Routledge.

Brown, T. A. (2015). Confirmatory Factor Analysis for Applied Research. New York, NY: The Guildford Press.

Carr, J., & Pauwels, A. (2005a). Boys and foreign language learning. Basingstoke: Palgrave Macmillan.

Carr, J., & Pauwels, A. (2005b). Boys and foreign language learning: Real boys don't do languages. Basingstoke: Palgrave Macmillan.

Chavez, 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, NJ: Lawrence Erlbaum Associates.

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Council of Europe (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume. Strasbourg: Council of Europe Publishing. Available at www.coe.int/lang-cefr

Csizér, K., & 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. http://doi.org/10.1093/applin/amn025

Csizér, K., & 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. doi.org/10.2167/jmmd557.0

Csizér, K., & Kormos, J. (2009). Learning experiences, selves and motivated learning behaviour: A comparative analysis of structural models for Hungarian secondary and university learners of English. In Z. Dörnyei & E. Ushioda (Eds.), Motivation, language identity and the L2 self (pp. 98–119). Bristol: Multilingual Matters.

Davis-Kean, P. E. (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. doi.org/10.1037/0893-3200.19.2.294

Dörnyei, Z. (2005). The psychology of the language learner: Individual differences in second language acquisition. Lawrence Erlbaum Associates.

Dörnyei, Z., & Ushioda, E. (Eds.). (2009). Motivation, langauge identity and the L2 self. Bristol: Multilingual Matters.

Dunn, K. & Iwaniec, J. (2022). Exploring the relationship between second language learning motivation and proficiency: A latent profiling approach. Studies in Second Language Acquisition. https://doi.org/10.1017/S0272263121000759

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

Eccles, J. S. (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. doi.org/10.1111/j.1471-6402.1994.tb01049.x

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

Erling, E. J., & 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 R. M.

Jimenez Catalan (Ed.), Gender perspectives on vocabulary in foreign and second languages (pp. 93–116). Basingstoke: Palgrave Macmillan.

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

Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second language acquisition. Canadian Journal of Behavioural Psychology, 13, 266–72.

Gardner, R. C., & Lambert, W. E. (1972). Motivational variables in second-language acquisition. In R. C. Gardner & W. E. Lambert (Eds.), Attitudes and motivation in second-language learning. Rowley, Mass.: 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., & Cliffordson, C. (2013). Motivation, gender, and possible selves. Language Learning, 63(2), 271–295. doi.org/10.1111/lang.12009

Hsieh, P. P.-H., & Kang, H.-S. (2010). Attribution and self-efficacy and their interrelationship in the Korean EFL context. Language Learning, 60(3), 606–627. http://doi.org/10.1111/j.1467-9922.2010.00570.x

Hsieh, P. P.-H., & Schallert, D. L. (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. http://doi.org/10.1016/j.cedpsych.2008.01.003

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

Iwaniec, J. (2014). Motivation of pupils from southern Poland to learn English. System, 45(1), 67–78. doi.org/10.1016/j.system.2014.05.003

Iwaniec, J. (2015). Motivation to learn English of Polish gymnasium pupils. Lancaster University. Retrieved from http://opus.bath.ac.uk/49838/

Iwaniec, (2019). Language learning motivation and gender: The case of Poland. International Journal of Applied Linguistics, 29(1). https://doi.org/10.1111/ijal.12251

Iwaniec, Janina & Halbach, Ana. (2021). Teachers' views on their methodology and their profiles: in search of the possible reasons for the levelling effect of CLIL. Journal of Multilingual and Multicultural Development. 10.1080/01434632.2021.1946548.

Iwaniec, J., & Ullakonoja, R. (2016). Polish and Finnish teenagers' motivation to learn English: The role of context. European Journal of Applied Linguistics, 0(0). http://doi.org/10.1515/eujal-2015-0034

Jimenez Catalan, R. M. (2010). Gender tendencies in lexical acquisition and use. In R. M. Jimenez Catalan (Ed.), Gender perspectives on vocabulary in foreign and second languages (pp. 117–138). Basingstoke: Palgrave Macmillan.

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

Klebanov, P. K., Brooks-Gunn, J., & Duncan, G. J. (1994). Does neighborhood and family poverty affect mothers' parenting, mental health, and social support. Journal of Marriage and Family, 56(2), 441–455. Retrieved from http://www.jstor.org/stable/353111

Kormos, J., & 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. http://doi.org/10.1016/j.system.2013.03.006

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

Kyriacou, C., & Zhu, D. (2008). Shanghai pupils' motivation towards learning English and the perceived influence of important others. Educational Studies, 34(2), 97–104. 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), 997–1023. 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. http://doi.org/10.1080/01434632.2012.697467

Marconi, G., Campos Cascales, C., Covacevich, C., and Halgreen, Tue (2020) What matters for language learning? The questionnaire framework for the PISA 2025 foreign language assessment. OECD Education Working Paper No. 234

Marsh, H. W. (1990). Self Description Questionnaire-I (SDQI) [Database record]. APA PsycTests.https://doi.org/10.1037/t01843-000

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. British Journal of Educational Psychology, 81, 59–77. http://doi.org/10.1348/000709910X503501

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

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

Muthén, L. K., & Muthén, B. O. (1998-2017). Mplus user's guide. (8th ed.). Los Angeles, CA: Muthén & Muthén.

Norton, B., & Pavlenko, A. (2004). Gender and English language learners: Challenges and possibilities. In B. Norton & A. Pavlenko (Eds.), Gender and English language learners (pp. 1–12). Virginia: TESOL.

O'Sullivan, B. (2011). Language testing. In J. Simpson (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, C. J. (2011). Language testing and validation. In B. O'Sullivan (Ed.) Language testing: theory & practice (pp.13–32). Oxford: Palgrave.

O'Sullivan, B., Dunlea, J., Spiby, R., Westbrook, C., and Dunn, K. (2020). Aptis General Technical Manual Version 2.2. Available online at: https://www.britishcouncil.org/exam/aptis/research/publications/technical/general-technical-manual-version-2-2 (accessed March 2023).

OECD. (2013). PISA 2012 results: Excellence through equity: Giving every student the chance to succeed (Volume II). OECD Publishing. 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/foreing 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 J. Sunderland (Ed.), Exploring gender: Questions and implications for English language education (pp. 140–147). Prentice Hall.

Pfenninger, S. E., & Singleton, D. (2016). Affect trumps age: A person- in-context relational view of age and motivation in SLA. http://doi.org/10.1177/0267658315624476

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78. http://doi.org/10.1037110003-066X.55.1.68

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

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistics. Psychometrika, 75, 243. http://doi.org/doi:10.1007/s11336-009-9135-y

Shavelson, R., Hubner, J.,&Stanton, G. (1976). Self-concept: Validation of construct interpretations. Review of Educational Research, 46, 407–441. https://doi.org/10.3102/00346543046003407.

Shepherd, E., & Ainsworth, V. (2017). English impact: An evaluation of English language usage capability: Madrid, Spain. British Council. https://www.britishcouncil.es/sites/default/files/british-council-english-impact-report-madrid-web-opt.pdf

Shepherd, E., & Ainsworth, V. (2018a). English impact: An evaluation of English language usage capability: Bogotá, Colombia. British Council. https://www.britishcouncil.org/sites/default/files/h203_01_english_impact_report_colombia_aw4.pdf

Shepherd, E., & Ainsworth, V. (2018b). English impact: An evaluation of English language usage capability: Uzbekistan. British Council. https://www.britishcouncil.org/sites/default/files/h203_01_english_impact_report_sri_lanka_aw5.pdf

Shohamy, E. (1998). Critical language testing and beyond. Studies in Educational Evaluation. 24 (4)

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. Sex Roles, 50(3/4), 241–252. http://doi.org/10.1023/B:SERS.0000015555.40976.e6

Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self-system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei & E. Ushioda (Eds.), Motivation, language identity and the L2 self (pp. 66–98). Bristol: Multilingual Matters.

Taylor, F., & Marsden, E. J. (2014). Choosing to study foreign languages in England : An experimental intervention. The Modern Language Journal, 98(4), 902–920. http://doi.org/10.1111/modl.12146

Ushioda, E. (2009). A person-in-context relational view of emergent motivation, self and identity. In Z. Dörnyei & E. Ushioda (Eds.), Motivation , language identity and the L2 self (pp. 215–228). Bristol.

Weir, C. J. (2005). Language Testing and Validation: an evidenced-based approach

Yashima, T. (2000). Orientations and motivation in foreign language learning: A study of Japanese college students. JACET Bulletin, 31, 121–133.

APPENDIX A – CEFR GLOBAL SCALE: COMMON REFERENCE LEVELS

Table A.1: CEFR Global Scale: Common Reference Levels[†]

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

[†]Retrieved from: https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

APPENDIX B - OVERVIEW OF THE APTIS FOR TEENS TEST STRUCTURE

Table B.1: Structure of the Teens Core component

Test	Part	Skill focus	Items/ Part	Level	Tasks/ level	Items/ Task	Task focus	Task description	Response format
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.	3-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.	5 target words. Select the best match for each from a bank of 10 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.	5 sentences, each with a 1-word gap. Select the best word to complete each from a bank of 10 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.	5 sentences, each with a 1-word gap. Select the best word to complete each from a bank of 10 options.
					1	5	Definition (vocabulary breadth)	Matching words to definitions.	5 definitions. Select the word defined from a bank of 10 options.
				B2	1	5	Collocation (vocabulary depth)	Word matching: match the word which is most commonly used with a word targeted from the appropriate vocabulary level.	5 target words. Select the best match for each from a bank of 10 options.

Table B.2: Structure of the Teens Reading component

Test	Part	Skill focus	Items/ Part	Level	Tasks/ level	Items/ Task	Task focus	Task description	Response format
Reading 25 items	1	Sentence level meaning	5	A1	1	5	Sentence level meaning (Careful, local reading)	Gap fill. A short text with 5 gaps. Filling each gap only requires comprehension of the sentence containing the gap. Text-level comprehension is not required.	3-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 6 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 4 short paragraphs giving information about 4 people's opinions on different topics and identify which of the four people could say certain statements.	7 gaps in a short text. Select the best word to fill each gap from a bank of 9 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.	7 paragraphs forming a long text. Select the most appropriate heading for each paragraph from a bank of 8 options.

Table B.3: Structure of the Teens Listening component

Test	Skill focus	Item/ Part	Level	Tasks/ level	Items/ Task	Task focus	Task description	Response format
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.).	3-option multiple choice. Only the target is mentioned in the text.
	Identifying specific, factual information	5	A2	5	1	Monologues & Dialogues	Q&A about listening text. Listen to short monologues and conversations to identify specific pieces of information (numbers, names, places, times, etc.).	3-option multiple choice. Lexical overlap between distractors and words in the input text.
	Identifying specific, factual information	5	B1	5	1	Monologues & 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.	3-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 & 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.	3-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.

Table B.4: Structure of the Teens speaking component

Test	Part	Skill focus	Level	Task description	Channel of input / prompts	Time to plan	Time for response	Rating criteria
Speaking	1	Giving personal information	A1/A2	Candidate responds to 3 questions on personal topics. The candidate records his/her 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: 1) grammatical range and accuracy 2) lexical range and accuracy 3) pronunciation 4) fluency 5) cohesion and coherence.
	2	Describing, expressing opinions, providing reasons and explanations	B1	The candidate responds to 3 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	
	3	Describing, comparing and contrasting, providing reasons and explanations	B1	The candidate responds to 2 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	
	4	Integrating ideas on a topic into a long turn presentation. Giving and justifying opinions, advantages and disadvantages	B2	The candidate plans a longer turn 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	2 minutes for the entire response	

Table B.5: Structure of the Teens Writing component

Test	Part	Skill focus	Level I	Task description	Channel of input / prompts	Expected output	Rating criteria
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 9 clearly marked categories (name, date of birth, etc.). There are 9 gaps in the form to be filled.	9 short gaps filled by 1–2 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): 1) task completion 2) grammatical range and accuracy 3) lexical range and accuracy 4) cohesion and coherence 5) punctuation and spelling.
	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 1. 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 3 separate 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 1 and 2.	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	
	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	

APPENDIX C – FINAL CFA MEASUREMENT MODEL ESTIMATES

Table C.1: Standardized model results, STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
INTOR BY				
INTOR1	0.470	0.028	17.011	0.000
INTOR2	0.712	0.019	36.906	0.000
INTOR3	0.685	0.022	30.613	0.000
INTOR4	0.666	0.022	30.322	0.000
PAR BY				
PAR1	0.563	0.025	22.499	0.000
PAR2	0.547	0.026	20.749	0.000
PAR3	0.751	0.020	36.751	0.000
PAR4	0.734	0.020	36.278	0.000
SELF BY				
SELF1	0.591	0.024	24.735	0.000
SELF2	0.705	0.022	32.260	0.000
SELF3	0.768	0.017	46.299	0.000
SELF4	0.693	0.021	33.569	0.000
OUGHT BY				
01	0.608	0.023	25.995	0.000
02	0.611	0.027	22.902	0.000
03	0.487	0.028	17.195	0.000
EX BY				
EX2	0.609	0.021	28.439	0.000
EX3	0.717	0.021	34.573	0.000
EX4	0.701	0.019	36.363	0.000
MB BY				
MB1	0.617	0.025	25.033	0.000
MB2	0.636	0.024	26.212	0.000
MB3	0.622	0.024	26.202	0.000

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
INSTR BY				
INSTR1	0.748	0.020	37.549	0.000
INSTR2	0.713	0.020	35.849	0.000
INSTR3	0.706	0.023	31.098	0.000
IDEAL BY				
I1	0.502	0.026	18.984	0.000
I2	0.585	0.023	25.100	0.000
I3	0.669	0.023	29.179	0.000
I4	0.667	0.020	33.142	0.000
PAR WITH				
INTOR	0.740	0.026	28.433	0.000
SELF WITH				
INTOR	0.463	0.031	14.987	0.000
PAR	0.402	0.031	12.886	0.000
OUGHT WITH				
INTOR	0.814	0.029	27.986	0.000
PAR	0.853	0.029	29.261	0.000
SELF	0.486	0.035	13.849	0.000
EX WITH				
INTOR	0.645	0.031	21.118	0.000
PAR	0.665	0.030	22.199	0.000
SELF	0.775	0.023	33.026	0.000
OUGHT	0.649	0.035	18.515	0.000
MB WITH				
INTOR	0.708	0.036	19.792	0.000
PAR	0.791	0.033	24.176	0.000
SELF	0.612	0.033	18.490	0.000
OUGHT	0.760	0.040	19.199	0.000
EX	0.922	0.029	31.968	0.000

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
INSTR WITH				
INTOR	0.819	0.023	36.395	0.000
PAR	0.852	0.025	34.173	0.000
SELF	0.429	0.031	13.868	0.000
OUGHT	0.799	0.033	24.320	0.000
EX	0.683	0.030	21.577	0.000
MB	0.774	0.034	22.891	0.000
IDEAL WITH				
INTOR	0.714	0.027	26.273	0.000
PAR	0.640	0.030	21.645	0.000
SELF	0.700	0.028	24.707	0.000
OUGHT	0.734	0.033	22.450	0.000
EX	0.763	0.029	26.197	0.000
MB	0.717	0.035	20.765	0.000
INSTR	0.576	0.031	18.323	0.000
MB3 WITH				
MB1	0.299	0.035	8.494	0.000
PAR 1 WITH				
PAR2	0.221	0.035	6.296	0.000
Intercepts				
INTOR1	3.736	0.092	40.763	0.000
I1	3.368	0.080	42.290	0.000
PAR	4.108	0.103	39.848	0.000
PAR2	4.401	0.115	38.227	0.000
SELF1	3.235	0.073	44.515	0.000
O1	3.841	0.090	42.736	0.000
INTOR2	4.266	0.115	37.104	0.000
I2	3.417	0.083	41.334	0.000
MB1	4.097	0.098	41.668	0.000
O2	3.954	0.102	38.872	0.000
INSTR1	4.810	0.152	31.600	0.000
INSTR2	4.602	0.120	38.202	0.000

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Intercepts				
MB3	4.080	0.104	39.298	0.000
SELF2	3.045	0.066	45.930	0.000
EX2	3.737	0.088	42.269	0.000
INSTR3	4.500	0.122	36.772	0.000
SELF3	3.240	0.071	45.425	0.000
INTOR3	4.000	0.098	40.970	0.000
EX3	4.109	0.102	40.170	0.000
SELF4	3.304	0.073	45.568	0.000
PAR3	4.439	0.123	36.225	0.000
I3	3.849	0.094	41.017	0.000
O4	3.551	0.082	43.429	0.000
I4	3.294	0.074	44.791	0.000
MB2	4.022	0.095	42.489	0.000
INTOR4	4.343	0.110	39.495	0.000
EX4	4.073	0.095	43.085	0.000
PAR4	4.672	0.131	35.652	0.000
Variances				
INTOR	1.000	0.000	999.000	999.000
PAR	1.000	0.000	999.000	999.000
SELF	1.000	0.000	999.000	999.000
OUGHT	1.000	0.000	999.000	999.000
EX	1.000	0.000	999.000	999.000
MB	1.000	0.000	999.000	999.000
INSTR	1.000	0.000	999.000	999.000
IDEAL	1.000	0.000	999.000	999.000
Residual Variances				
INTOR1	0.779	0.026	29.990	0.000
I1	0.748	0.027	28.239	0.000
PAR1	0.683	0.028	24.255	0.000
PAR2	0.700	0.029	24.240	0.000
SELF1	0.651	0.028	23.089	0.000

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Residual Variances				
O1	0.631	0.028	22.213	0.000
INTOR2	0.493	0.027	17.951	0.000
I2	0.658	0.027	24.175	0.000
MB1	0.619	0.030	20.347	0.000
O2	0.626	0.033	19.185	0.000
INSTR1	0.441	0.030	14.816	0.000
INSTR2	0.491	0.028	17.316	0.000
MB3	0.613	0.030	20.719	0.000
SELF2	0.502	0.031	16.282	0.000
EX2	0.629	0.026	24.069	0.000
INSTR3	0.501	0.032	15.616	0.000
SELF3	0.410	0.026	16.062	0.000
INTOR3	0.531	0.031	17.319	0.000
EX3	0.486	0.030	16.373	0.000
SELF4	0.520	0.029	18.210	0.000
PAR3	0.436	0.031	14.195	0.000
I3	0.552	0.031	17.973	0.000
O4	0.763	0.028	27.726	0.000
I4	0.554	0.027	20.623	0.000
MB2	0.596	0.031	19.309	0.000
INTOR4	0.556	0.029	19.007	0.000
EX4	0.509	0.027	18.838	0.000
PAR4	0.461	0.030	15.520	0.000

APPENDIX D – ADDITIONAL OUTPUT FROM CFA

Table D.1: Error covariances included in the measurement model

Covariance ref	Question ref	Question details	Reduction in adjusted Chi-square value	Significance
1	I1	I imagine myself speaking English fluently.	52.157	*p<.001*
	I2	I imagine myself comfortably reading in English on the internet.		
2	INSTR1	I need English for my future career.	57.820	*p<.001*
	INSTR2	The things I want to do in the future require me to use English.		
3	PAR1	My parents think I need to know English to be well-educated.	29.237	*p<.001*
	PAR2	My parents have stressed the importance English will have for		
4	SELF1	I usually get good marks in English.	29.754	*p<.001*
	SELF3	I have always done well in English.		
5	SELF1	I usually get good marks in English.	12.370	*p<.001*
	SELF4	Studying English comes easy to me.		
6	O2	Studying English is important to me because other people will respect me more if I have knowledge of English.	9.552	*p<.001*
	O3	Studying English is important to me because an educated person is supposed to be able to speak English.		
7	O1	I consider learning English important because the people I respect think that I should do it.	15.790	*p<.001*
	O2	Studying English is important to me because other people will respect me more if I have knowledge of English.		

Table D.2: T-tests results for gender differences

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
IDEAL fscore for gender model	Equal variances assumed	.057	.811	2.221	1316	.026	.07196	.03239	.00841	.13551
	Equal variances not assumed			2.218	1235.237	.027	.07196	.03245	.00830	.13562
PAR fscore for gender model	Equal variances assumed	.150	.698	.764	1316	.445	.02729	.03574	-.04282	.09740
	Equal variances not assumed			.762	1231.159	.446	.02729	.03583	-.04300	.09758
SELF fscore for gender model	Equal variances assumed	.621	.431	4.008	1316	.000	.15493	.03866	.07909	.23076
	Equal variances not assumed			3.987	1217.170	.000	.15493	.03886	.07869	.23117
EX fscore for gender model	Equal variances assumed	.590	.442	3.112	1316	.002	.11987	.03851	.04432	.19542
	Equal variances not assumed			3.120	1253.937	.002	.11987	.03842	.04449	.19525
INSTRU fscore for gender model	Equal variances assumed	1.538	.215	1.652	1316	.099	.07743	.04688	-.01453	.16939
	Equal variances not assumed			1.639	1201.848	.102	.07743	.04726	-.01528	.17015
MOTBEH fscore for gender	Equal variances assumed	1.114	.292	3.277	1316	.001	.15500	.04730	.06221	.24780
	Equal variances not assumed			3.261	1219.968	.001	.15500	.04753	.06176	.24824
ought fscore for gender	Equal variances assumed	1.113	.292	1.306	1316	.192	.05167	.03956	-.02594	.12928
	Equal variances not assumed			1.298	1211.286	.195	.05167	.03982	-.02644	.12979

Table D.3: Freeing up latent covariances between gender groups in the constrained model: Calculations of adjusted Chi-square value (Satorra & Bentler, 2010)

Constrained model:		MOST CONSTRAINED MODEL	Add: PARENT with IDEAL relaxed	Add: INSTRU with PAR relaxed	Add: INSTRU with IDEAL relaxed
degrees of freedom	*d0*	615	614	613	612
scaling correction factor	*c0*	1.30	1.30	1.30	1.30
MLM chi-square value	*T0*	1305.44	1298.96	1293.52	1285.02
More relaxed model:		PARENT with IDEAL relaxed	Add: INSTRU with PAR relaxed	Add: INSTRU with IDEAL relaxed	Add: OUGHT with INSTRU
degrees of freedom	*d1*	614	613	612	611
scaling correction factor	*c1*	1.30	1.30	1.30	1.30
MLM chi-square value	*T1*	1298.96	1293.52	1285.02	1280.44
Calculations					
Diff test scaling correction	*cd*	1.49	1.49	1.73	1.48
Adjusted chi-sq value	*Calc*	8.83	7.47	11.97	6.34
Adjusted chi-sq value	*TRd*	**5.94**	**5.03**	**6.92**	**4.27**