

Developing GSE Job Profiles

Part 2: Creating and validating a model of
language proficiency for job roles

February 2024

Authors:

Dr. Rebecca Dew, Catherine Doyle

Contributors:

Dr. Jooyoung Lee, Claire Chettiar,
Mike Mayor, Dr. William Bonk,
Lauren Miller



Contents

| | |
|---|-----------|
| Introduction | 5 |
| PART 1: Creating the model | 6 |
| 1.1 Background | 6 |
| About the Global Scale of English | 6 |
| The problem to solve | 6 |
| 1.2 Methodology | 7 |
| Occupation ontologies | 7 |
| Determining baseline occupation requirements | 8 |
| Re-scaling attribute scores to the GSE scale | 9 |
| <i>Adjustment for communication requirements</i> | 10 |
| <i>Adjustment for role seniority</i> | 11 |
| <i>Adjustment for occupation-specific regulations</i> | 12 |
| GSE level ranges | 12 |
| 1.3 Findings and discussion | 14 |
| Qualitative review | 14 |
| Relative proficiency across occupation categories | 14 |
| GSE level by language skill | 17 |
| PART 2: External industry validation study | 20 |
| 2.1 Introduction | 20 |
| 2.2 Methodology | 20 |
| The standardised set of responses | 21 |
| Industry panel participants (Raters) | 21 |

| | |
|--|-----------|
| Familiarisation | 22 |
| The rating process | 22 |
| Methodology for determining an optimal cut score | 23 |
| Post-grading panel meeting..... | 23 |
| 2.3 Findings and discussion..... | 24 |
| Insights into the industry..... | 24 |
| Rater statistics..... | 24 |
| Optimal cut scores on Versant scale | 26 |
| Comparing Versant Benchmarking Kit findings to GSE Job Profiles | 27 |
| Conclusions and next steps..... | 30 |
| References | 31 |
| Appendices | 32 |
| Appendix 1: Job roles..... | 32 |
| Appendix 2: Rater statistics for each role/skill in the study | 33 |

Introduction

GSE Job Profiles answers the question “What level of English language is needed to perform a particular job?” It is the first-of-its-kind model for speakers of English as a foreign, second, or additional language who need to use English for work, either in an English-speaking country or for an international company who use English as a lingua franca to communicate with partners and customers in different countries around the world. GSE Job Profiles recommends a level of proficiency at which an employee could comfortably perform most tasks in a given occupation. It is presented as a range on the Global Scale of English (GSE), reflecting the fact that there will be differences in levels of seniority, experiences and circumstances within the same job role. The ranges serve as guidelines around which companies can set their own benchmarks depending on the local context and requirements.

GSE Job Profiles is of benefit to:

- Employers wishing to establish English language requirements as part of the recruitment process
- Employers wishing to set English language learning goals as part of their retention and upskilling strategy for existing staff
- Learners and future job applicants looking to understand the English language requirements for their chosen profession
- Teachers and trainers planning to create and deliver English language courses for specific professions

This paper is divided into two sections; in the first we describe how the model was developed, based on attribute data in O*NET (USDOL/ETA) and Pearson Workforce ontology. The second part of the paper describes how we validated the model with panels of industry experts.

PART 1: Creating the model

1.1 Background

About the Global Scale of English

The GSE is a standardised English proficiency scale from 10 to 90 that is psychometrically aligned to the Common European Framework of Reference for Languages (de Jong, Mayor and Hayes, 2016). Sets of GSE Learning Objectives for Adult, Academic, Professional, Young Learners and Pre-Primary have been developed to describe learner proficiency at each point on the scale.

Examples

- *Can discuss details of a work-related task (e.g. deadlines, specifications).*
GSE 59, Professional, Speaking.
- *Can write very short, simple sentences about their feelings.*
GSE 32, Adult, Writing.

The work to develop the GSE Learning Objectives builds upon the research carried out by Brian North and the Council of Europe in creating the CEFR (North, 2000), incorporating and extending the CEFR Can Do statements. The GSE Learning Objectives have been developed by Pearson over a number of years in collaboration with more than 6,000 teachers, ELT authors and language experts from around the world (de Jong, Mayor & Hayes, 2016).

The problem to solve

English is indisputably the *lingua franca* of international business (Roshid, Webb, Chowdry, 2022). This has been the case for many years as a result of the globalisation of the world's economies and English has only increased in importance in a post-pandemic world where remote working has become more commonplace, facilitated by the advance of technologies, enabling employers to recruit staff anywhere in the world. More than ever, employers are building teams across multiple locations who speak multiple first languages. In those situations, English is often the company's official language of communication.

Against this backdrop, companies still face the same core challenges they have always had; how to recruit and retain the best talent. The only difference now is that an increasing number of organisations are having to factor English language proficiency into their recruitment and retention strategies.

Some of the common problems we hear from employers are:

- We spend so much time sifting applications
- It takes a lot of resources to check applicants' English at the interview stage
- Even applicants with an English qualification are not always able to communicate easily in English in the workplace
- Our biggest challenge is staff engagement
- Younger generations are focused as much, if not more, on professional development and job satisfaction as they are on salary

We believe that businesses would benefit from having a better understanding of the English language proficiency needed for different job roles, as a way to speed up the recruitment process and to build out an organisational language policy that identifies English language training needs to support the development of their employees.

Our problem statement for this research was: *“What level of English is needed to successfully perform tasks in a particular role?”*

1.2 Methodology

Occupation ontologies

Occupation titles and the requirements of each role vary by organisation and location, and this makes planning and insights into an organisations workforce difficult. Occupation ontologies provide standardised names and structures for each unique occupation, enabling the compilation of occupation-based data for analysis and workforce planning. English proficiency requirements for a role form part of such data.

The Occupational Information Network (O*NET) ontology is a widely-used, publicly-available ontology sponsored by the U.S. Department of Labor. O*NET contains data on approximately 1000 occupations specifically relevant to the US economy. The O*NET ontology provides scores for 244 attributes which represent occupation-related competencies. These attributes cover such things as knowledge, skills, abilities and work activities. The scores for these attributes indicate the importance of the attribute and level of competency required for each occupation, based on surveys of US workers.

The Pearson Workforce (PW) ontology was used as the underlying ontology for the English Language model described in this report. It expands on the O*NET ontology to cover approximately 1400 occupations. All occupations in the PW ontology have also been mapped to the same set of attributes as O*NET. In the PW, attribute scores have been converted to a single, continuous scale from 0-1.

Determining baseline occupation requirements

A baseline English language proficiency for each occupation was determined from the PW occupation attribute scores, drawing on those attributes directly related to reading, writing, speaking and listening (Table 1).

Table 1: Attributes used to determine the baseline English requirements per occupation. All attributes directly related to each language skill were chosen.

| Language Skill | Attribute names |
|----------------|--|
| Reading | Reading Comprehension, Written Comprehension |
| Writing | Writing, Written Expression |
| Speaking | Oral Expression, Speaking, Speech Clarity |
| Listening | Active Listening, Oral Comprehension, Speech Recognition |

Attribute scores reflect the level of skill that a job requires for that attribute; a higher score indicating the job requires a higher level of skill. In the original O*NET attribute framework there are multiple attributes which can be directly related to each language skill, for example 'Written Expression' and 'Writing' both refer to writing skills. Occupations generally have different scores for each of these attributes. For example, Foreign Affairs Policy Officers have a higher score for 'Written Expression' (0.62) than 'Writing' (0.54), while Telecommunications Policy Officers have a higher score for 'Writing' (0.53) than 'Written Expression' (0.49). There is no discernable pattern indicating which attribute best represents each language skill. One possible explanation is that there are differences in the way in which 'Written Expression' and 'Writing' are interpreted by people in different occupations. The assumption was made that the highest attribute score for each language skill reflects the minimum requirement for that occupation so the maximum attribute score for each language skill was taken. For example, Foreign Affairs Policy Officers would have a writing score of 0.62 from its highest scoring writing-related attribute 'Written Expression' (vs 0.54 for 'Writing') while Telecommunications Policy Officers would have a writing score of 0.53

from the 'Writing' attribute (vs 0.49 for 'Written Expression'). The resulting scores gave the relative skill required in each occupation for reading, writing, speaking and listening on a scale of 0-1. The average of these four individual language skills gave the overall baseline attribute score for each occupation.

As the O*NET data underlying these scores is collected from surveys in the USA, these scores were assumed to reflect the language requirements for a first language (L1) speaker of English working in an English-speaking country. An external industry validation study on the results of this model was carried out to assess this assumption (see Part 2) and ensure that the recommended proficiency levels were applicable to speakers of English as a second language (L2) working in a variety of countries and contexts.

Re-scaling attribute scores to the GSE scale

The next step was to re-scale the language attribute scores to the GSE scale, adjusting the score range of 0-1 to an appropriate GSE level. GSE levels range from 10, being a complete beginner, to 90, being able to understand and communicate to a high level of proficiency in almost any context. In a professional work environment, most employees whose first language is not English would not be expected to have English proficiency at either extreme of this scale. For this reason, the attribute scores were linearly scaled to a narrower GSE range of 20-80.

Language requirements in the workplace are related to many other factors, such as the nature of communication tasks and the potential impact of communication errors. In order to accurately reflect workplace English requirements, therefore, additional O*NET data was included to adjust the baseline occupation scores. The aim was to capture information on additional non-language attributes indicating the importance of communication in certain roles that may have been missing in the attributes directly related to language skills. Occupations that were identified as having a higher communication requirement were placed on a slightly expanded scale of 20-85 (see next section). This achieved the effect of slightly tweaking the relative order of closely scored occupations while preserving the overall data structure.

Adjustment for communication requirements

As mentioned above, the baseline attribute scores (Table 1) did not fully reflect the importance of communication and social skills in certain occupations. O*NET includes many non-language attributes related to these communication and social skills in the workplace. A subset of further attributes was chosen to reflect these additional communication requirements for each occupation (Table 2). Some of these attributes clearly relate to specific language skills, for example, the attribute 'Telephone' is related to speaking and listening. Others imply language skills, such as 'Work with Work Group or Team'.

Where any of the additional attributes for reading, writing or speaking/listening were greater than 0.5, the occupation was regarded as requiring some expertise in that language skill. Other attributes could not be aligned to a specific language skill but indicate a general requirement for communication e.g., 'Deal With External Customers' or 'Developing and Building Teams' could involve all four language skills depending on the occupational context. Occupations with a score greater than 0.5 for any of the general communication attributes were regarded as having a requirement for general communication skills.

Where an occupation had a requirement for both general communication and expertise in one or more language skill, the score for that specific skill was re-scaled. For example, an occupation that scored above 0.5 for 'Deal With External Customers' would have a general communication requirement. If this occupation also scored high for 'Telephone', the speaking/listening score of that occupation would be re-scaled.

The attribute scores were rescaled on a scale of 20–85. As a result, the scores for occupations in which communication is important were slightly higher than the original scores calculated on the scale of 20–80.

Table 2: Attributes used to determine the communication requirements per occupation. Each attribute was assigned to a category based on whether it indicated a general communication requirement or if it indicated a communication requirement using a specific language skill.

| Language Usage Requirements | Attribute names |
|-----------------------------|--|
| General Communication | 'Deal With External Customers', 'Work With Work Group or Team', 'Contact With Others', 'Coordinate or Lead Others', 'Communicating with Supervisors, Peers, or Subordinates', 'Communicating with People Outside the Organization', 'Performing for or Working Directly with the Public', 'Coordinating Work and Activities of Others', 'Developing and Building Teams', 'Training and Teaching Others', 'Guiding, Directing, and Motivating Subordinates', 'Coaching and Developing Others' |
| Reading | 'Electronic Mail', 'Letters and Memos' |
| Writing | 'Electronic Mail, Letters and Memos', 'Documenting/Recording Information' |
| Speaking/Listening | 'Providing Consultation and Advice to Others', 'Telephone', 'Face to Face Discussions', 'Resolving Conflicts and Negotiating with Others' |

Adjustment for role seniority

Initial assessment of the baseline attribute scores surprisingly showed that many managerial roles had language scores lower than the related individual contributor (IC) roles. Broadly across the workplace there is a general assumption that skill levels increase with the seniority of the role, including language skills. Furthermore, managerial and leadership roles often have a greater proportion of tasks that require business-critical communication. There are of course exceptions to this, such as in customer service occupations where non-managerial roles have the most interaction with external customers. However, to allow the managerial roles to reflect the likely higher English requirements, these roles were rescaled to an expanded range. A range of 20–90 was used so that the resulting attribute scores for managerial roles were scaled to slightly higher GSE values than their equivalent IC roles, including IC roles with communication specialisations on the 20–85 scale.

Adjustment for occupation-specific regulations

For some occupations there are strict minimum English requirements enforced by regulatory bodies or industry standards. Data on occupation-specific English proficiency requirements were collected from research studies and regulatory bodies. These were used as a reference point for considering the model outputs and comparison of the variability in proficiency required. In most cases these data cannot be used directly to adjust occupation scores as the scores reported may be generalised for the industry as a whole, vary over time due to changes in regulations (e.g. visa entrance scores), be country-specific, or the level may be set based on additional factors outside of just English proficiency requirements (e.g. entrance exam levels may be set to adjust the pool of potential applicants).

One exception to this is military roles which have international requirements outlined by NATO for the level of English required in specific posts. Based on the NATO STANAG 6001 language requirements (BILC, 2019), a minimum GSE of 43 was enforced for all military roles, allowing these occupations to be accurately scaled to international standards. The higher end of the scale was set based on the communication requirements of the role as outlined above.

GSE level ranges

According to our project hypothesis, the GSE level for each job should represent a level of English whereby an individual can comfortably perform most tasks in a 'typical' role. The language competency needed for a specific role, however, can vary considerably. The language capability needed can be impacted by the range of tasks required in that role, the skills across the broader team, the company's use of additional services (e.g. paying for editorial services to review written work), the employees' broader communication skills, company culture etc. To reflect this variability, the GSE levels were presented as a range rather than a single number. Companies can use the GSE ranges as guidance while retaining the flexibility to set their own cut-scores for the specificities of a role.

Seniority level is a common factor which can add variability to the level of English needed in a role. In the PW ontology, each occupation is broken down into different seniority levels: junior, associate (mid), senior and expert (Figure 1). The PW has attribute scores that have been adjusted to represent the expected capabilities in these roles. These seniority-adjusted attribute scores were used to establish the size of the GSE range.

Looking across each occupation, there was, on average, a four-point difference between each seniority level; from junior to mid, mid to senior and senior to expert. Therefore, four points were added either side of the mid-seniority GSE levels to represent the variability expected in a role between junior and senior levels.

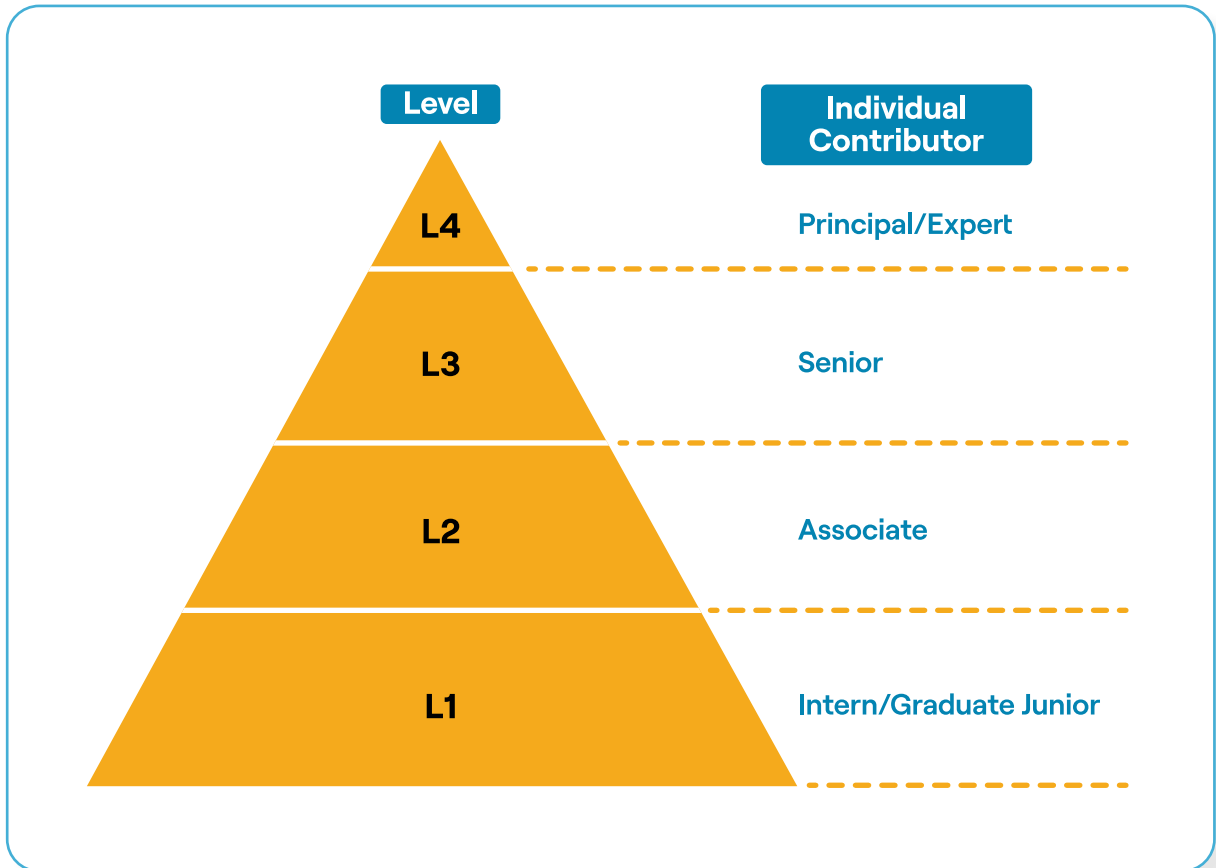


Figure 1: Occupation levels with junior positions at the bottom in level one and experts at the top in level four.

1.3 Findings and discussion

Qualitative review

The results for all occupations were qualitatively reviewed by a subject matter expert who specialises in teaching Business English and is familiar with the GSE. They scrutinised the data and flagged 33 anomalous occupations. Sommeliers, for example, were flagged for having a low GSE score of 30–38, equivalent to Food Preparation Workers. Sommeliers generally work in more prestigious establishments which sets a higher expectation on the language proficiency required. This type of consideration is not captured in the O*NET attribute data.

Some anomalies are to be expected as the underlying O*NET data was not originally collected with English proficiency in mind. Survey data is also subject to large variance in responses, particularly as people in different occupations may interpret questions on the importance of specific attributes to their work in very different ways.

Flagged occupations were suppressed from the results, leaving 1372 occupations.

Relative proficiency across occupation categories

Overall, the relative order of English language proficiency reflects what would commonly be expected across different occupation categories (Figure 2). The legal, scientific and medical sectors have the highest requirements on average. Construction & mining had the lowest requirements on average, as a result of a large number of more manual occupations being levelled around GSE 30 and despite having some occupations that require 70+.

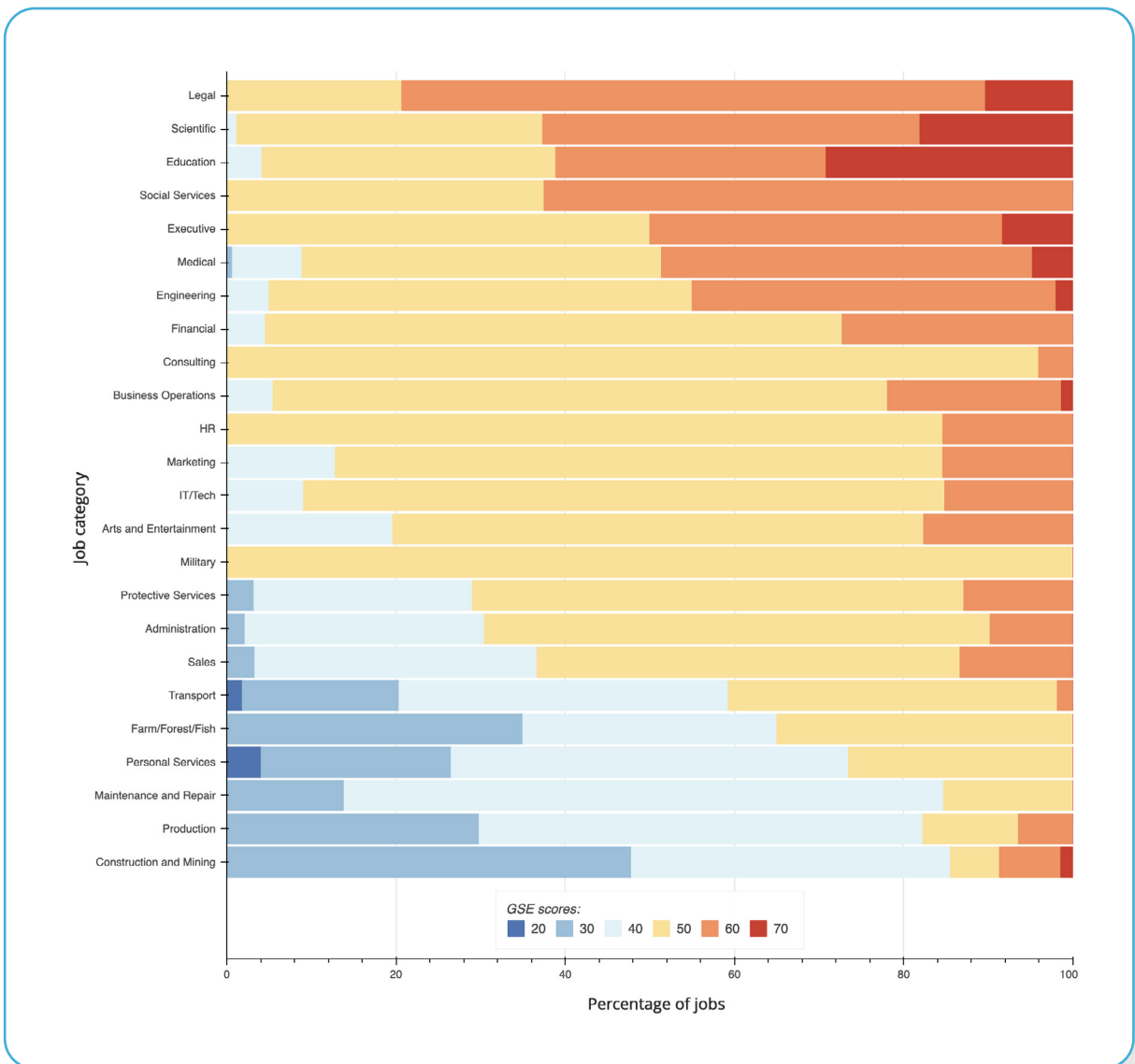


Figure 2: Proportional distribution of occupations across the GSE scale for each category. Categories are ordered by average GSE level. The Scientific, Education and Medical sectors have the most jobs at the highest end of the scale (70+).

The occupations in each sector also follow a relative order that would be expected (Figure 3). For example, in the administration sector, ‘Order Fillers’, who have a manual and routine set of tasks have a range of GSE 35–43.

Order Fillers:

“Fill customers’ mail and telephone orders from stored merchandise in accordance with specifications on sales slips or order forms. Duties include computing prices of items, completing order receipts, keeping records of out-going orders, and requisitioning additional materials, supplies, and equipment.”

‘Switchboard Operators’ and ‘Human Resources Assistants’ fall between 44–52, reflecting the higher English requirements in these roles for communication with customers and personnel.

Switchboard Operators:

“Operate telephone business systems equipment or switchboards to relay incoming, outgoing, and interoffice calls. May supply information to callers and record messages.”

Human Resources Assistants:

“Compile and keep personnel records. Record data for each employee, such as address, weekly earnings, absences, amount of sales or production, supervisory reports, and date of and reason for termination. May prepare reports for employment records, file employment records, or search employee files and furnish information to authorised persons.”

‘Foreign Affairs Policy Officers’ are ranked higher with GSE levels from 62–70, as they are required to interpret policies and write reports in a technical area of expertise, and therefore require stronger English skills.

Foreign Affairs Policy Officers:

“Analyse foreign affairs policies and operations, and write reports outlining the analyses. Communicate with parties who benefit from the findings, and act as advisor in the development, implementation, or reporting on foreign policy. May also perform administrative duties in the department, such as assisting with problems concerning passports and visas.”

Other sectors also follow an expected ordering of English language proficiency, reflecting the communication requirements, technical expertise and level of responsibility in these roles.

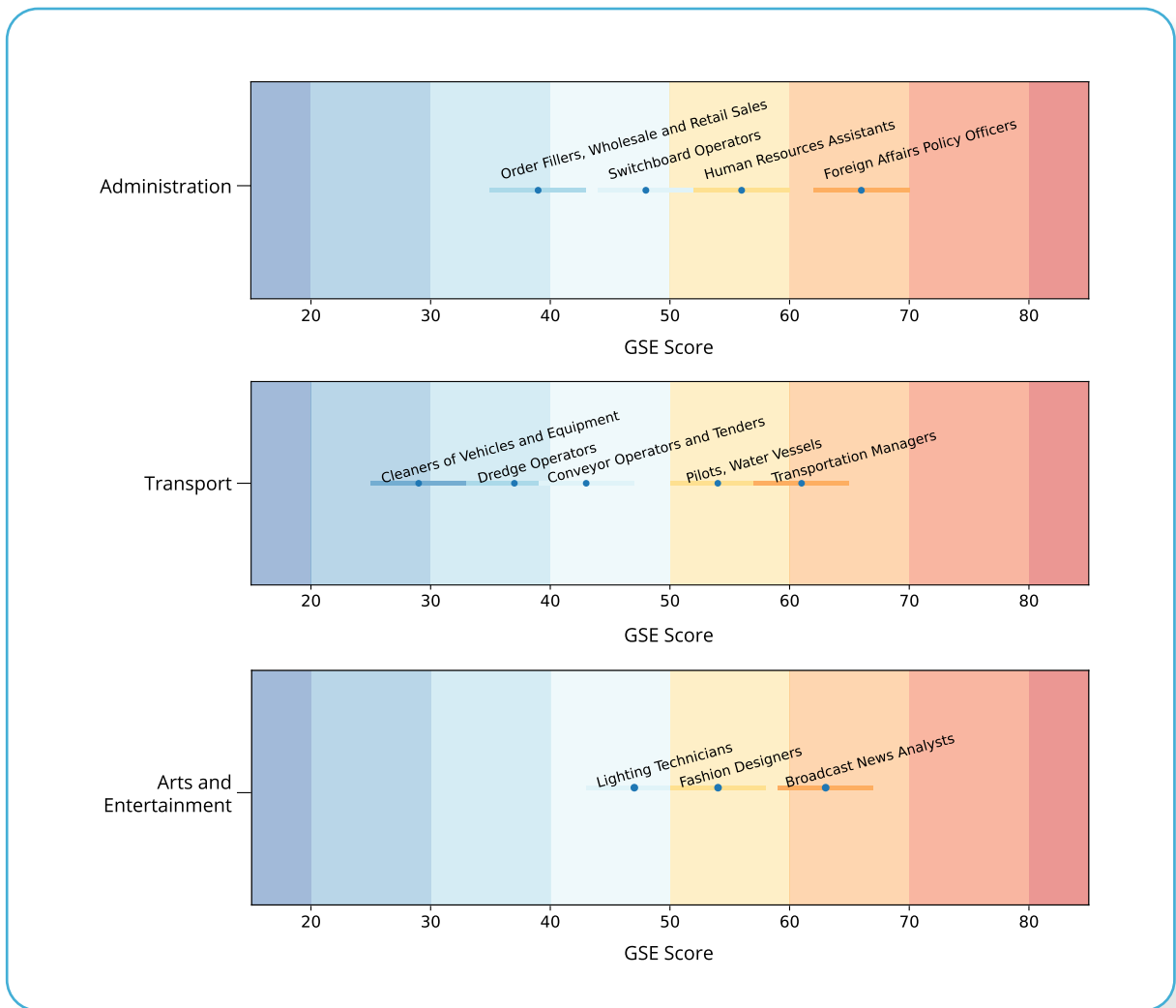


Figure 3: Examples of occupations falling within each GSE band. The dot indicates the average GSE level for the occupation.

GSE level by language skill

The overall GSE levels can be further broken down into the individual language skills for reading, writing, speaking and listening. The occupations with the highest English proficiency requirements in each language skill are comprised of occupations with specialisations in those skills, such as reading for Chief Editors (Figure 4).

Listening and speaking skills are commonly linked to people-oriented roles, as we associate these skills with the additional social skills required for this type of communication. On the GSE scale, the highest levels reflect the understanding and usage of technical or field-specialised language. Therefore, jobs such as biophysicists have high GSE levels for listening and speaking, while jobs with a more social focus that require less technical language do not require such high GSE levels (Figure 5).

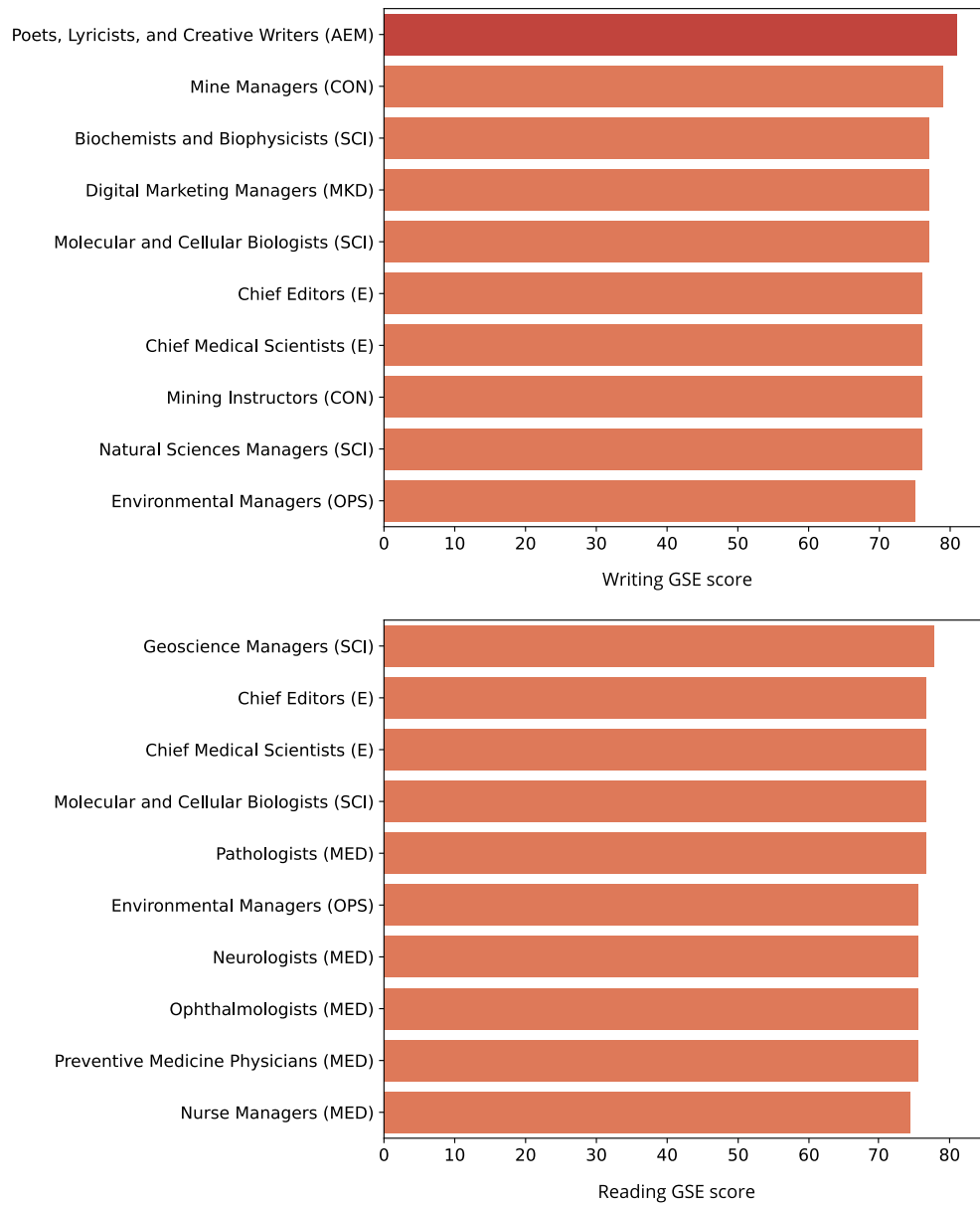


Figure 4: Top 10 occupations by reading and writing. Poets, Lyricists, Creative Writers and Digital Marketing Managers, have high GSE writing levels as might be expected. Mine Managers also have high GSE levels for writing due to the technical information they are required to communicate and the importance of clear communication.

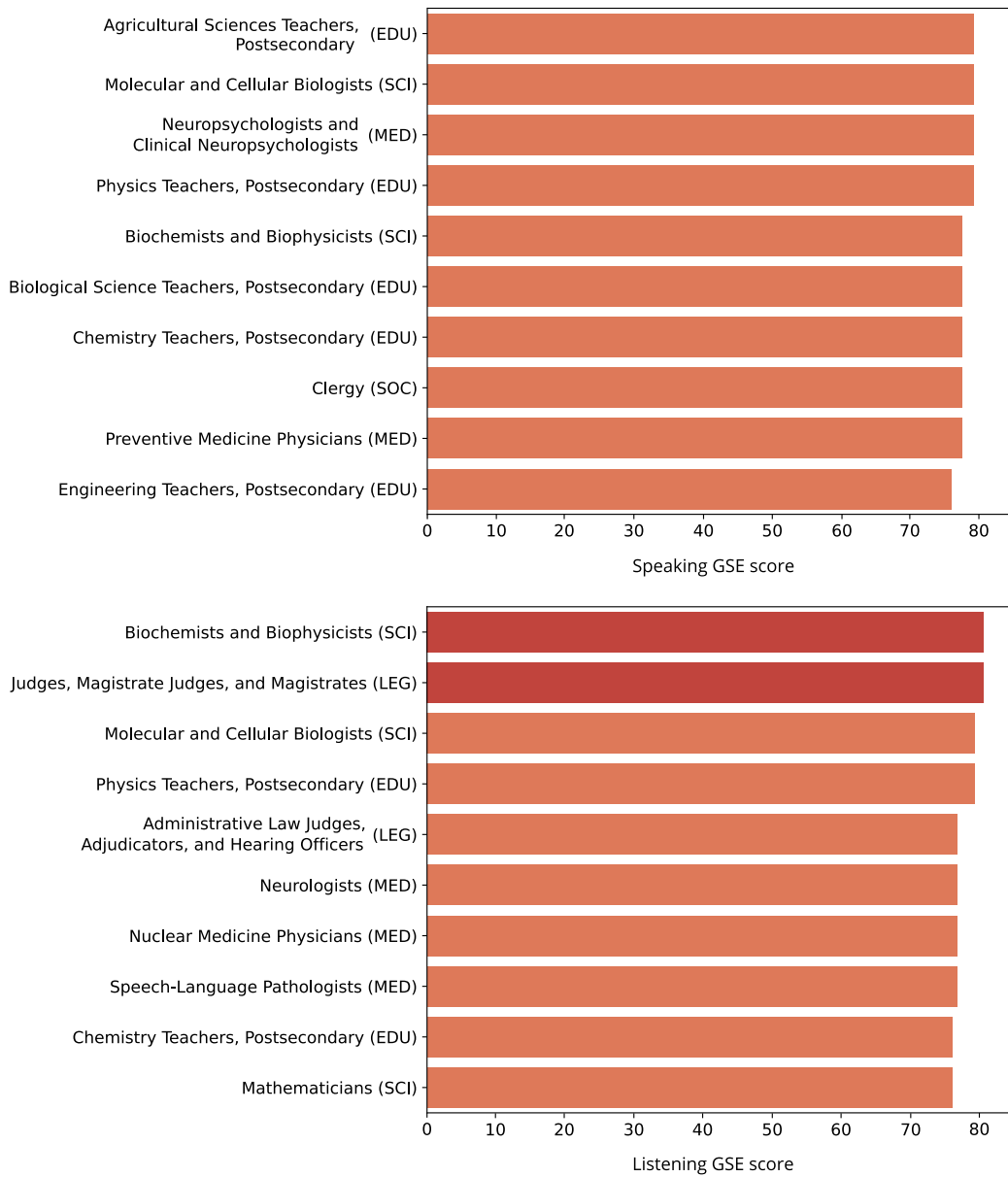


Figure 5: Top 10 occupations by speaking and listening. Many education occupations appear in the top 10 by speaking and listening as do medical occupations.

PART 2: External industry validation study

2.1 Introduction

The purpose of this study was to validate the data model outlined in Part 1 and ensure that the recommended ranges were in line with the expectations and existing practices of those working in the industry. We set out to discover:

- What are the levels of English proficiency that industry experts would expect for different roles?
- Are there any differences in English language requirements for job roles in different countries/contexts? Would the GSE Job Profile recommendations be acceptable globally?
- What could we learn from industry experts that we could use to further refine the model?

2.2 Methodology

We set up industry expert panels for three industries; hospitality, finance and customer service. The study employed a benchmarking methodology using test-taker responses sampled from Pearson Versant Writing Test (VWT) for writing, and Versant English Test (VET) for speaking (for more information about these tests, visit [Versant, our tests | Pearson Languages](#)). The panelists were asked to read or listen to test-takers' responses and judge whether they demonstrated adequate written/spoken English ability for a specific role. With this data, we established speaking and writing cut scores on the GSE for two different jobs within each industry. This result was then compared with the GSE ranges established for the same roles within GSE Job Profiles, using the data model outlined in Section 1 of this report.

Table 3: Job roles selected for the study in each industry. The two job roles were selected to reflect different levels of English proficiency, see appendix for job descriptions for each role.

| Industry | Job Role 1 | Job Role 2 |
|------------------|--|-------------------|
| Hospitality | Dining Room and Cafeteria Attendant | Lodging Manager |
| Finance | Insurance Claims and Policy Processing Clerk | Actuarial Manager |
| Customer Service | Customer Service Representatives | Helpdesk Manager |

The standardised set of responses

The study used a standardised set of responses that had been brought together as a 'Versant Benchmarking Kit' and used in the past for other similar studies. 57 VET and 51 VWT test-takers were identified for use in the benchmarking kit and accent; it also represented an even spread of language ability across the full 20–80 Versant scoring scale. Two responses per test-taker were selected and then screened to verify that they were representative of their overall VET/VWT score. The 102 written responses and 114 spoken responses were randomised for presentation to the panelists. In this way, the Versant Benchmarking Kit provides a standardised and well-established dataset of spoken and written English to be analysed by non-language specialists when identifying the level of English proficiency required for a particular job.

Industry panel participants (Raters)

Each panel had between 6–11 members representing professional bodies and employers from around the world.

Table 4: Industry Panel demographics

| Panel | Number of members | Countries represented |
|------------------|-------------------|---------------------------------------|
| Hospitality | 11 | UK, US |
| Finance | 9 | Brazil, India, UK |
| Customer Service | 6 | Colombia, Brazil, UK, US, Philippines |

The panelists were recruited through various channels, including callouts from industry bodies, personal connections and membership of existing Pearson industry panels for other projects. Potential panelists were required to complete an application form, from which the most suitable participants were selected.

All the panelists held senior posts within their industry and were involved in making or influencing recruitment decisions (i.e., HR manager, L&D manager, operations manager). Many of them had experience in recruiting people whose first language was not English. They all spoke a high-level of English themselves (L1 English or L2 speaker at GSE 75–90/CEFR C1/C2). The panels were as diverse as possible in terms of gender, nationality and experience.

The panelists committed about 5 hours of their time over a 6-week period and were paid for their participation.

Familiarisation

The panelists were given a short pre-read task to familiarise themselves with the project along with descriptions of the two job roles to be analysed (Table 3). They met virtually as a group, across two or three alternative sessions, in order to accommodate availability and time zones. In the first part of the virtual meeting, the panelists discussed two questions:

- *What consideration do you have, if any, for language skills when you're recruiting?*
- *What kinds of English language skills are needed for different job roles in your industry?*

In the second part of this meeting, the group were asked to consider 'what is the minimum standard of English would you expect for these two roles?' They were guided by the chair to consider the different skills (e.g., Reading, Writing, Speaking, and Listening) and subskills (e.g. pronunciation, fluency, accuracy, grammar, vocabulary) that each of the two roles would require to carry out their tasks in English, then come to some agreement as a group on what level they would expect and why. For the customer service panel, the group also co-created a persona for the two roles. This standardisation exercise aimed to ensure that the group had a common understanding of the scope of the two roles that they would bear in mind whilst rating (whilst recognising that the remit of the role might differ from organisation to organisation) and ensure the panelists were clear on what level of English they would each expect to see, in order to ensure consistency in their own judgements.

The rating process

Each of the panelists (referred to from hereon as *raters*) was assigned job roles to grade; they would grade both speaking AND writing skills for one job role and either speaking OR writing for the other. The assignment of roles was carried out using an overlapping model to ensure that a different combination of raters was assessing each job role.

Each of the raters was given access to the Versant Benchmarking Kit in an online tool. They were presented with spoken or written responses. For each response, the raters were instructed to answer the question, "*Does this test taker demonstrate the minimum standard of English expected for this role?*", clicking 0 for no and a 6 for yes.

An average overall rating for each test-taker was produced. Those test-takers who received an average rating below 3 were considered *Unacceptable*, and test-takers with an average rating above 3 were considered Acceptable. An overall average rating of 3.0 meant the test-taker received the same number of 0 and 6 ratings from the raters. Such test-takers were excluded from the analysis because the rater ratings were not conclusive as to whether their English level was suitable for that job.

Methodology for determining an optimal cut score

A cut-score is point on the scale which represents minimum level of English required for that role, establishing a consensus across the different raters. As with all activities using human judgment and opinion, a certain amount of noise and uncertainty in the data is to be expected. Although it is not always possible to identify a cut score that eliminates disagreements completely, an optimal cut score represents the best possible solution.

A statistical analysis was used to identify the Versant cut score that most agrees with the rater decisions of Acceptable/Unacceptable for the sets of test-takers. All possible cut scores were tested, and the best one presented as the *Optimal Versant Threshold* (see Results). If Versant scores are successfully discriminating between acceptable and unacceptable candidates, the percentage of agreement between the selected raters will be high.

The Versant cut scores (on the Versant Scale) were then converted to values on the GSE using an established transformation function in order that a direct comparison to GSE Job Profiles could be made.

Post-grading panel meeting

Once the cut scores had been established, the panels re-convened in virtual meetings. The numerical findings were presented to the panelists for discussion along with explanations of those scores using both sample responses and descriptions of score meanings. The panelists were given the chance to further discuss or dispute the cut scores and give feedback on the grading process. In the final exercise, the panel discussed and rank-ordered 10-15 job roles within their industry according to the level of English they thought the roles required, using the two previously graded jobs as benchmarks.

2.3 Findings and discussion

Insights into the industry

All three industry panels agreed that English language is an important skill required for almost all jobs in their industry, although the level of attention that it gets during recruitment varies according to both the country and the individual company. The following high-level points were observed:

- **Hospitality:** There were nuances here between the type of establishment recruiting for the roles and the expectations for English, with roles like maids and waiting staff needing higher levels of English for higher-end hotels and restaurants. Additionally, the pandemic (as well as Brexit in the UK) has also meant that there is far more demand than supply, which has led to less stringent requirements for language skills.
- **Finance:** This is a high-stakes environment governed by regulatory groups; employees must be able to communicate complex content to the public, and this comes with a risk of litigation. The jargon and terminology are highly specialised. As many finance jobs require a university degree, English-language ability is often assumed rather than overtly requested, particularly in English-speaking countries.
- **Customer service:** Listening and Speaking skills are generally valued higher than reading and writing, particularly with the advent of Chat-GPT and similar technologies. Customer Service roles are often outsourced to places like India and the Philippines, where English is a second language rather than a foreign language. On-shore clients, however, still have high-expectations when it comes to English proficiency, which means English ability is very much part of the recruitment process.

Rater statistics

The individual rater's performance was evaluated for each of the roles and skills. The 'Rater Average' represents the average rating that a particular rater assigned across all the test-takers in the sample. Rater averages were compared to understand how much variability there was in the criterion levels among individual raters, and how strict or lenient raters were. The rater average is not a reflection of whether a rater did a "good" or "bad" job with the judgments – only whether they were relatively strict or generous in their judgments.

A rater whose average was closer to 0 judged more candidates as *Unacceptable*; such a rater is considered relatively “strict” regarding the level of English required. In other words, the rater felt that only fairly high levels of English ability were acceptable. A rater whose average was closer to 6, on the other hand, rated more candidates as *Acceptable*; such a rater is considered relatively “generous.” That is, the rater found more than half of the candidates in the dataset Acceptable.

Table 5: Individual rater statistics for Lodging Manager - Writing Set

| Rater Number | Rater Average |
|--------------|---------------|
| 1 | 1.11 |
| 2 | 2.11 |
| 3 | 1.81 |
| 4 | 1.82 |
| 5 | 2.06 |
| 6 | 1.53 |
| 7 | 2.71 |
| 8 | 3.17 |

The table above shows an example of this for the Writing sample for the role of Lodging Manager. The analysis shows that there is only a small amount of variation among raters, indicating that they closely agree on the optimal English level required. The optimal cut score in this example is a consensus value, with the more lenient (e.g., rater 8) and stricter (e.g., rater 1) raters cancelling each other out.

The rater statistics across all the roles are shown in Appendix 2. Raters for the two job roles in Hospitality showed little variation in their ratings. This general consistency suggests that the raters were well-informed about the rating criteria and shared a very similar understanding of the proficiency level required for those roles. On the contrary, raters for the job roles in Finance demonstrated greater variability in their judgments. This inconsistency might be due to the more heterogeneous nature of this group of panelists (e.g., different L1 backgrounds, working in different areas within the field of Finance), the heterogeneity of the field itself, or insufficient training. For the Customer Service industry, it is hard to draw a solid conclusion about the performance of raters because only a small number (4–5) participated in the rating exercise.

Optimal cut scores on Versant scale

For each set of data, the optimal cut score was produced after testing all possible cut scores. In some cases, more than one score was equally acceptable, and therefore a range is given.

As an example, results of the *Lodging Manager (Writing)* analysis are displayed in Figure 6 below. The horizontal line at 56 represents the optimal VWT cut score produced, meaning anyone scoring at or above 56 should be considered acceptable. Data points shown as large circles represent candidates for whom VWT scores and raters' ratings aligned, and those shown as Xs represent candidates who were classified differently by the two methods. Small circles in the chart represent cases where rater average ratings for candidates were exactly 3.0.

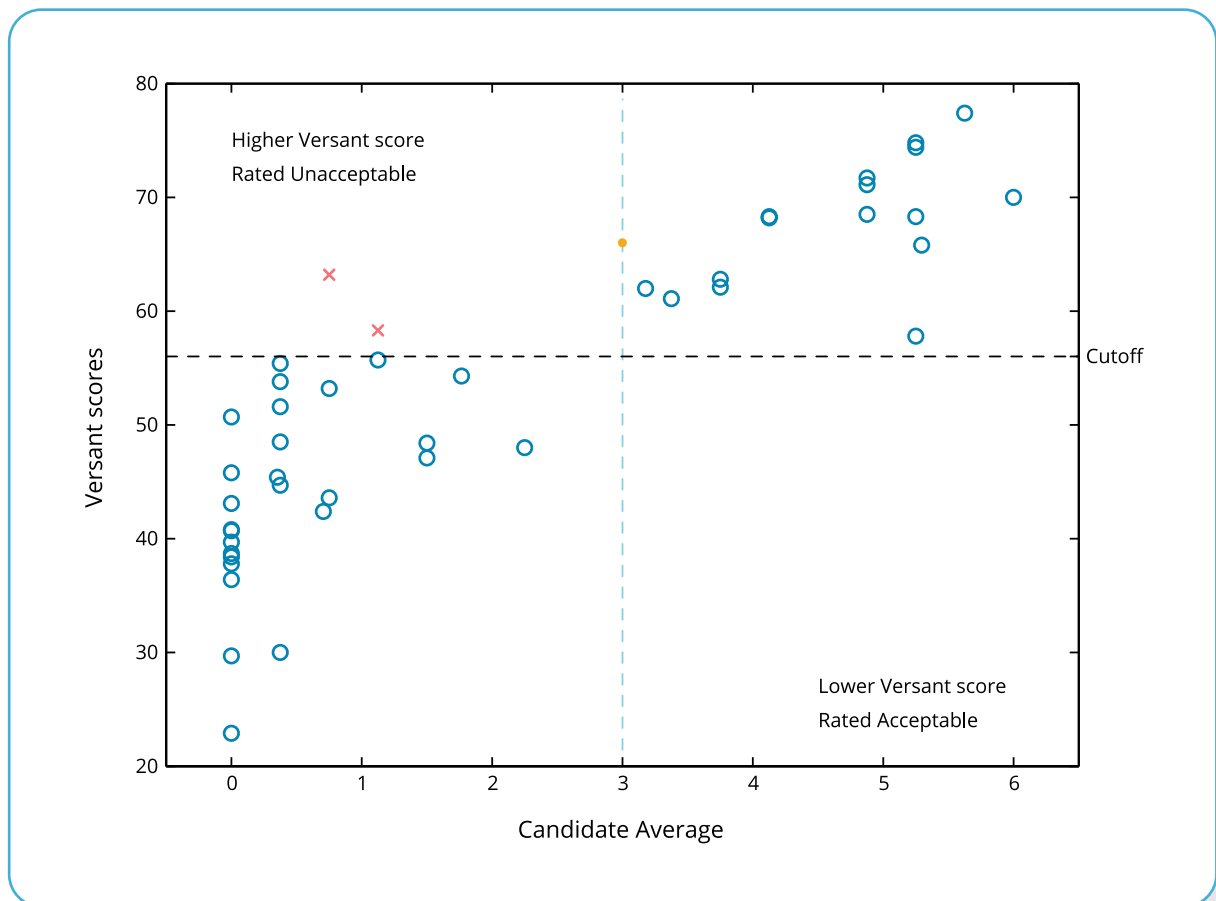


Figure 6: Average candidate grade based on Lodging Manager (Writing) rater judgments, VWT scores, and agreement between the two methods (n = 51).

This analysis demonstrates that, for 48 out of 50 test-takers (96%), a VWT cut score of 56 is in agreement with the raters' acceptability judgments. One case was eliminated from the analysis because raters gave them average ratings of exactly 3.0 and was therefore inconclusive. As shown in Figure 6, only two disagreements (marked as x) were found using the calculated optimal cut score of 56. The VWT cut score of 56 identified in this analysis is therefore a highly accurate one that aligns very closely with the panel's perceptions of required English ability.

The agreement rate between the two approaches (cut score and industry panel judgment) ranges from 89% to 96% in all job roles and skills. This suggests that an identified cut score for each role/skill accurately represents the raters' perceptions of required English ability. Thus, new test-takers obtaining a test score at or above the cut-scores are nearly certain to have the English proficiency level required for those roles.

Comparing Versant Benchmarking Kit findings to GSE Job Profiles

The table below shows the GSE cut scores established in the Versant Benchmarking Kit studies compared with the ranges established in the GSE Job Profile model. The colour-coding shows where the GSE optimal cut score from the Versant Benchmarking Kit results falls within (Blue), close to (Yellow) or outside (Pink) the GSE Job profile range.

Table 6: Comparing Versant Benchmark Kit results to GSE Job Profiles levels.

| Role | Skill | Number of Raters in analysis | Optimal cut-score from Versant Benchmark Study on the GSE (CEFR in brackets) | GSE Job Profiles level (CEFR in brackets) |
|------------------------|----------|------------------------------|--|---|
| Cafeteria Attendant | Speaking | 8 | 35 (A2) | 28-36 (A2) |
| | Writing | 8 | 40 (A2+) | 23-31 (A1) |
| Lodging Manager | Speaking | 9 | 48 (B1) | 53-61 (B1+) |
| | Writing | 8 | 62 (B2) | 59-67 (B2) |
| Insurance Claim Clerk | Speaking | 7 | 48 (B1) | 44-52 (B1) |
| | Writing | 7 | 62 (B2) | 44-52 (B1) |
| Actuarial Manager | Speaking | 6 | 59 (B2) | 59-67 (B2) |
| | Writing | 7 | 62/66 (B2) | 61-69 (B2) |
| Customer Service Agent | Speaking | 4 | 45-46 (B1) | 47-55 (B1/B1+) |
| | Writing | 5 | 56-62 (B1+/B2) | 47-55 (B1/B1+) |
| Helpdesk Manager | Speaking | 4 | 49-53 (B1/B1+) | 46-54 (B1/B1+) |
| | Writing | 5 | 56-58 (B1+) | 57-65 (B2) |

For Hospitality, the speaking level given by the panel for the cafeteria attendant role was at the top of the recommended range in GSE Job Profiles and the writing level came out significantly higher. The converse was true for the Lodging Manager where the Versant results came out 5 points lower for speaking ability but did match very well on writing. In the follow-up meeting with the panelists, the group all agreed that a cafeteria attendant would not need a high level of writing and they would be happy with the ranges proposed by GSE Job Profiles. The higher-end hotels would still prefer a cafeteria attendant with a higher level of speaking proficiency since they might come into contact with the public. Similarly, they all agreed that the higher speaking level given by GSE Job Profiles for Lodging Manager was more suitable than the Versant cut score, since this role would require managerial-level communication to co-ordinate staff as well as dealing with customer enquiries and complaints.

For Finance, the results for Actuarial Manager matched well with GSE Job Profiles. The panel examined the role of the insurance claim clerk again, with a focus on the writing level where there is a considerable discrepancy between the benchmarking results (62/B2) and the GSE Job Profiles range (44-52/B1).

Insurance claim clerk:

“Process applications for, changes to, reinstatement of, and cancellation of insurance policies. Duties include reviewing insurance applications to ensure that all questions have been answered, compiling data on insurance policy changes, changing policy records to conform to insured party’s specifications, compiling data on lapsed insurance policies to determine automatic reinstatement according to company policies, canceling insurance policies as requested by agents, and verifying the accuracy of insurance company records.”

The discussion focused on how much of the role would be templated and routine, and whether the writing needs could be considered “simple” and “on familiar topics” (i.e. B1) or if there was an amount of independent, detailed report writing that might be required on a wider range of topics (i.e. B2). The panel were satisfied that as this is a clerk role, the writing ability given in GSE Job Profiles would be sufficient.

For Customer Service, the Help-Desk Manager role results matched well with GSE Job Profiles data. The Customer Service Agent role results, whilst only just outside the ranges of GSE Job Profiles, warranted further panel discussion.

Customer Service Representative:

“Interact with customers to provide basic or scripted information in response to routine inquiries about products and services. May handle and resolve general complaints. Excludes individuals whose duties are primarily installation, sales, repair, and technical support.”

As with the Insurance Claim Clerk role, the discussion focused on the level of scripted and templated language that would be used in the role, as well as the expectations of the hiring company and their location. One panelist from the Philippines is already using a GSE score of 58 for recruitment purposes and felt that the GSE Job Profiles range (47-55) was more suitable than the Versant Benchmarking Kit cut-score (45-46) but still not high enough for their context.

Conclusions and next steps

Overall, there was a good level of alignment at the skill-level within the two datasets, strengthening our belief that the individual job roles can be given a recommended English proficiency level, and moreover that the suggested GSE ranges are acceptable to employers globally. The validation results indicate that the ranges suggested in GSE Job Profiles not only capture the base English skills required, but also the interconnected communication requirements of specific roles. In the future, our validation data and overall validation strategy will allow us to expand the model to further occupations and explore models for specific locations.

In the light of this new research, we are also revisiting the earlier work carried out (Gordon, Hayes, Mayor & Buckland, 2018) and re-aligning work activities specified by O*NET for specific job roles to GSE Learning Objectives. This detailed role mapping enables teachers and trainers who are creating and/or delivering English language courses for specific professions to have a better understanding of how to target their instruction and create specific learning pathways.

References

Bureau for International Language Coordination (BILC, 2019). *NATO STANAG 6001, Ed. 5: Overview of language proficiency levels*. Retrieved 04.01.2024 from [BUREAU FOR INTERNATIONAL LANGUAGE COORDINATION \(natobilc.org\)](https://www.natobilc.org/)

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

De Jong, J.H.A.L., Mayor, M. and Hayes, C. (2016). *Developing Global Scale of English Learning Objectives aligned to the Common European Framework*. Available at [GSE-WhitePaper-Developing-LOs.pdf \(pearson.com\)](https://www.pearson.com/whitepaper/developing-LOs.pdf)

Gordon, M., Hayes, C., Mayor M., & Buckland, S. (2018) *Developing GSE Job Profiles: Interim Report and Initial Validation Study*. Available at [Developing GSE Job Profiles: Interim Report and Initial Validation Study \(pearson.com\)](https://www.pearson.com/developing-gse-job-profiles-interim-report-and-initial-validation-study)

National Center for O*NET Development. *O*NET Resource Center*. Retrieved January 4, 2024, from <https://www.onetcenter.org/>

North, B. (2000). *The Development of a Common Framework Scale of Language Proficiency*. New York: Peter Lang

Roshid, M. M., Webb, S., & Chowdhury, R. (2022). English as a Business Lingua Franca: A Discursive Analysis of Business E-Mails. *International Journal of Business Communication*, 59(1), 83-103. <https://doi.org/10.1177/2329488418808040>

Appendices

Appendix 1: Job roles

Table 7: Description of the Job roles selected for the study in each industry.

| Role | Short Description |
|---|--|
| Dining Room and Cafeteria Attendants | Facilitate food service. Clean tables; remove dirty dishes; replace soiled table linens; set tables; replenish supply of clean linens, silverware, glassware, and dishes; supply service bar with food; and serve items such as water, condiments, and coffee to patrons. |
| Lodging Managers | Plan, direct, or coordinate activities of an organisation or department that provides lodging and other accommodations. |
| Insurance Claims and Policy Processing Clerks | Process new insurance policies, modifications to existing policies, and claims forms. Obtain information from policyholders to verify the accuracy and completeness of information on claims forms, applications and related documents, and company records. Update existing policies and company records to reflect changes requested by policyholders and insurance company representatives. |
| Actuarial Managers | Manage a team of actuaries to analyse statistical data and construct probability tables to forecast risk and liability for the payment of future benefits. May ascertain insurance rates required and cash reserves necessary to ensure payment of future benefits. |
| Customer Service Representatives | Interact with customers to provide basic or scripted information in response to routine inquiries about products and services. May handle and resolve general complaints. Excludes individuals whose duties are primarily installation, sales, repair, and technical support. |
| Help Desk Managers | Supervise the help desk team ensuring that customers get the appropriate feedback and support. Monitor the delivery of technical support services to clients according to the customer service guidelines. Plan and organise user support actions and troubleshoot problems and issues. |

Appendix 2: Rater statistics for each role/skill in the study

Table 8: Individual rater statistics Cafeteria Attendant - Speaking

| | | | | | | | | |
|----------------------|------|------|------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Rater Average | 4.05 | 4.84 | 3.95 | 4.58 | 4.53 | 5.37 | 4.84 | 4.63 |

Table 9: Individual rater statistics Cafeteria Attendant - Writing

| | | | | | | | | |
|----------------------|------|------|------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Rater Average | 3.47 | 3.06 | 4.88 | 4.51 | 4.53 | 4.34 | 5.24 | 3.12 |

Table 10: Individual rater statistics Lodging Manager - Speaking

| | | | | | | | | | |
|----------------------|------|------|------|------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Rater Average | 2.26 | 1.47 | 2.58 | 2.85 | 2.79 | 3.05 | 0.42 | 2.95 | 3.42 |

Table 11: Individual rater statistics Lodging Manager - Writing

| | | | | | | | | |
|----------------------|------|------|------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Rater Average | 1.11 | 2.11 | 1.81 | 1.82 | 2.06 | 1.53 | 2.71 | 3.17 |

Table 12: Individual rater statistics Insurance Claims and Policy Processing Clerks - Speaking

| | | | | | | | |
|----------------------|------|------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Rater Average | 4.32 | 1.26 | 4.16 | 2.89 | 3.79 | 3.26 | 2.95 |

Table 13: Individual rater statistics Insurance Claims and Policy Processing Clerks - Writing

| | | | | | | | |
|----------------------|------|------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Rater Average | 4.59 | 2.94 | 1.24 | 3.76 | 2.29 | 1.12 | 0.69 |

Table 14: Individual rater statistics Actuarial Managers – Speaking

| | | | | | | |
|----------------------|------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 | 6 |
| Rater Average | 1.84 | 2.11 | 3.58 | 0.63 | 4.32 | 2.47 |

Table 15: Individual rater statistics Actuarial Managers – Writing

| | | | | | | | |
|----------------------|------|------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Rater Average | 1.47 | 0.76 | 2.82 | 2.71 | 2.29 | 1.53 | 2.59 |

Table 16: Individual rater statistics Customer Service Representatives – Speaking

| | | | | |
|----------------------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 |
| Rater Average | 4.05 | 3.74 | 2.79 | 3.58 |

Table 17: Individual rater statistics Customer Service Representatives – Writing

| | | | | | |
|----------------------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 |
| Rater Average | 1.92 | 2.65 | 4.12 | 1.94 | 3.29 |

Table 18: Individual rater statistics Help Desk Managers – Speaking

| | | | | | |
|----------------------|------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 | 5 |
| Rater Average | 3.63 | 3.79 | 1.95 | 2.63 | 1.58 |

Table 19: Individual rater statistics Help Desk Managers – Writing

| | | | | |
|----------------------|------|------|------|------|
| Rater Number | 1 | 2 | 3 | 4 |
| Rater Average | 3.06 | 1.96 | 3.18 | 2.47 |

Be yourself
in English.

