



## Research Report

# Alignment of Wall Street English levels to the Common European Framework and Global Scale of English

September 2019

**Simon Buckland**, MSc., Dip. TEFL  
**William Lorié**, PhD

# Author biographies

**Simon Buckland** is an independent educational consultant specialising in curriculum and assessment who has had a life-long involvement with technology-assisted language learning. He was previously Academic Director of Wall Street English, where he piloted a major large-scale quantitative project measuring learning and aligning the course to the Common European Framework. As Chief Course Developer, he pioneered the use of CEFR Learning Objectives as the basis for course content. From 2010 to 2017 he was the curriculum specialist on Pearson's Global Scale of English team, working to extend the CEFR descriptor set especially at lower (pre-A1) and higher levels. In this capacity he developed a set of CEFR-aligned grammar descriptors (published as the GSE Grammar Learning Objects). Simon was educated at Oxford University, where he took a B.A. in English, and at Sussex University, where he took an MSc in Artificial Intelligence, specialising in Intelligent Computer-Aided Learning, and is now an independent educational consultant specialising in curriculum and assessment.

**William Lorié**, Ph.D., is an assessment consultant who has held business development, senior scientist, and technical advisory positions at McGraw-Hill, ETS, Pearson and the World Bank. Dr. Lorié was awarded and directed a U.S. Department of Defense contract to develop and validate language proficiency tests for nine languages. He was awarded a grant through the U.S. Department of Education Institute for Education Sciences to develop a prototype conversation-based English language learning system. Dr. Lorié is currently principal at Capital Metrics, which provides assessment advisory services.

# Contents

<b>1. Executive summary</b> .....	4
<b>2. Background: The Common European Framework, the Global Scale of English, and Pearson Progress</b> .....	5
2.1. The Common European Framework .....	5
2.2. The Global Scale of English and Learning Objectives .....	6
2.3. Pearson Progress .....	7
<b>3. Background: Wall Street English levels</b> .....	8
<b>4. Purpose of the present project</b> .....	9
<b>5. Design and methodology</b> .....	10
<b>6. Procedure and data cleaning</b> .....	11
6.1 Selecting the students for the study .....	11
6.2 Taking the test and recording the results .....	11
6.3 Cleaning the data .....	12
<b>7. Analysis and results</b> .....	14
7.1 Overview .....	14
7.2 Analysis of data set 1 .....	14
7.3 Analysis of data set 2 .....	18
<b>8. Conclusions and next steps</b> .....	20
<b>References</b> .....	21

# 1 / Executive summary

The content of the instructional programme used by Wall Street English since 2016 is closely based on Pearson's Global Scale of English (GSE) Learning Objectives, which builds on and extends the Common European Framework or CEFR, the de facto global standard for English language testing instruction. The content of the Wall Street English instructional programme is therefore aligned with the most widely recognised global standards for English language learning.

Wall Street English uses this alignment between its own levels and the CEFR/GSE to determine prospective learners' targets in terms of the levels needed to reach their overall objectives for English communication. It was therefore necessary to verify the alignment in terms of output as well as input, by determining whether Wall Street English students generally attain the CEFR/GSE scores which would be predicted by their level within the Wall Street English instructional programme.

To achieve this, a six-skills medium-stakes test (Pearson Progress) was used. Progress, which reports an overall GSE score/CEFR level as well as separate scores for enabling skills such as Vocabulary, Grammar etc., was administered to a total of 4143 Wall Street English students in 107 learning centres in 11 territories. The resulting data set was filtered to remove students without a valid overall score, and the remaining 1615 records were analysed statistically to determine the degree of correlation between Wall Street English level and test score, and how close this was to Wall Street English's published alignment. A further subset of this data was created, composed of students with valid Progress scores in all enabling skills, and analysed following the same methodology to check the validity of the previous findings. This subset represents students who completed all areas of the test.

A high degree of correlation was found to exist between students' actual performance and their predicted performance according to their Wall Street English level, especially for the first ten levels. Wall Street English's published alignment of its own levels to CEFR and GSE was found to be accurate up to Level 12, and adequately accurate up to Level 14. Beyond Level 14 the data were insufficient to make an accurate determination.

## 2 / Background:

# The Common European Framework, the Global Scale of English, and Pearson Progress

### 2.1. The Common European Framework

The Common European Framework of Reference for Modern Languages (CEFR) (Council of Europe 2001) defines three broad classes of language ability: A - Basic User, B - Independent User, and C - Proficient User. Each of these is subdivided into two levels (1 and 2), giving six in all, from A1 to C2.

The CEFR defines language ability in functional or operational terms, not by reference to examinations or pre-existing standards - in other words, what someone is able to do at each level. The overall Can Do statements for the six levels are as follows:

Table 1: CEFR levels and definitions

Level	Name	Can Do definition
A1	Breakthrough	Can communicate in basic English with help from the listener
A2	Waystage	Can communicate in English within a limited range of contexts
B1	Threshold	Can communicate essential points and ideas in familiar contexts
B2	Vantage	Can use English effectively, with some fluency, in a range of contexts
C1	Effective Operational Proficiency	Able to use English fluently and flexibly in a wide range of contexts
C2	Mastery	Highly proficient - can use English very fluently, precisely and sensitively in most contexts

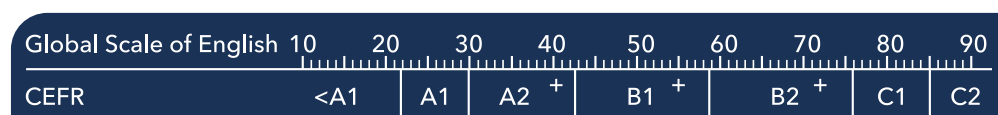
Each of these six levels in turn is defined by several Can Do statements covering the key skills of Reading, Writing, Listening and Speaking. The CEFR is empirically based on a research project carried out by Brian North (North 2000) in which experienced English teachers were asked to assess what tasks their learners could perform at various levels. The resulting findings were then analysed statistically and used to create the sets of Can Do statements which compose the CEFR.

The CEFR has become the de facto international standard for teaching and assessing English and is used as a reference scale in most published course materials and tests around the world.

## 2.2. The Global Scale of English and Learning Objectives

The Global Scale of English (GSE) (see Pearson 2018a) is a standardised granular scale developed by Pearson to support a more fine-grained understanding of English language proficiency than is possible with the CEFR. The GSE has been psychometrically aligned to the CEFR (De Jong, Mayor and Hayes 2016), and represents a series of finer gradations within the overall CEFR framework. The diagram below shows the relationship between the GSE and CEFR levels:

Figure 1: Global Scale of English and CEFR



The GSE is part of an 'ecosystem' composed of four main parts:

1. The scale itself
2. The GSE Learning Objectives
3. Course materials created using GSE
4. Assessment tools that report on GSE

The GSE Learning Objectives were developed to extend the CEFR and address its limitations, notably:

- a. The relatively small number of Can Do statements, leaving many gaps, especially at lower and higher levels,
- b. The very wide bands in the CEFR, often corresponding to many years of language study, and
- c. The fact that the CEFR starts at a level which is well above that of absolute beginners.

The GSE Learning Objectives include all the Can Do statements in the CEFR and add a complementary set of many hundreds of additional ones. These GSE Can Do statements, referring to a more granular scale, give teachers and learners access to a much more precise picture of English proficiency and language development. Moreover, the GSE scale starts at 10, well below A1 (which starts at 22), and therefore enables progress to be described and assessed at the most basic level.

## 2.3. Pearson Progress

The GSE 'ecosystem' referred to above includes several assessment tools, or tests, such as Pearson Test of English Academic (Pearson 2017) and Progress (2018b). Progress is a six-skills test which measures learners' English language proficiency on the Global Scale of English scale of 10 to 90 and hence also on the CEFR.

Progress combines computer-adaptive and non-adaptive parts in one test. It is a fully automatically scored test which includes productive speaking and writing skills, so ensuring that learners' proficiency levels and progress are measured accurately and efficiently. To ensure the most accurate results possible across the full range of abilities, Progress exists in six versions, covering different overlapping levels of ability. Learners are encouraged to take the test appropriate for their expected or notional level of English:

Table 2: Progress test levels and CEFR

<b>Progress</b>	<b>CEFR equivalent</b>
<b>15-30</b>	Below A1 to A1
<b>25-40</b>	Mid-A1 to A2
<b>35-50</b>	Mid-A2 to low B1
<b>45-60</b>	Low B1 to low B2
<b>55-70</b>	High B1 to mid-B2
<b>65-80</b>	Mid-B2 to C1

Progress reports an Overall score as well as six sub-scores: Reading, Writing, Speaking, Listening, Grammar and Vocabulary. If a student's score is out of range for the test they have taken (below the 'floor' or above the 'ceiling') Progress cannot compute a score with sufficient confidence, so it returns an advisory note instead: for example, 'Below A1' or 'Above A2'. Similarly, if not enough items have been completed, because for instance the test has been abandoned before the end, a 'low confidence' cut off is triggered, and no score is reported.

Progress has an internal consistency reliability coefficient of 0.91, and a test-retest reliability of 0.86. A correlation coefficient of 0.97 was observed during validation studies between overall scores given by expert human raters and those produced by Progress (See Pearson 2015). The standard error of measurement (SEM) of Progress is  $\pm 3$  GSE points.

# 3 / Background:

## Wall Street English levels

Wall Street English was an ‘early adopter’ of the CEFR: all courses since the 1980s were based on the Threshold Level syllabus (van Ek and Trim, 1979 & 1990) which was the principal precursor of the CEFR. Then in 2006-7, with the support of Cambridge ESOL, Wall Street Institute (as it was then called) carried out the world’s first globally based and fully empirical standard-setting study in English language learning (Buckland 2010). This study demonstrated that Wall Street Institute English language levels were aligned to the CEFR.

In 2012, when Wall Street English was a subsidiary of Pearson English, the decision was taken to create a completely new course, to be called the Wall Street English Core Course. To ensure compatibility with the most widely recognised international standards for measuring learner progression, the syllabus was based closely on the GSE Learning Objectives (see above). Each exercise and module was keyed to a Learning Objective or group of Learning Objectives, and each assessment within the course was designed to test the learners’ mastery of one or more Learning Objectives.

Each level of the Core Course is divided into four units of mixed multimedia and classroom-based instruction, making a total of 80 units as shown below:

Table 3: Wall Street English stages, levels and units with related CEFR and GSE levels

<b>Stages</b>	<b>Survival</b>	<b>Waystage</b>	<b>Upper Waystage</b>	<b>Threshold</b>	<b>Milestone</b>	<b>Mastery</b>
<b>Levels</b>	1-2	3-5	6-9	10-13	14-17	18-20
<b>Units</b>	1-8	9-20	21-36	37-52	53-68	69-80
<b>CEFR</b>	Below A1	A1	A2	B1	B2	C1
<b>GSE</b>	10-21	21-29	29-42	42-58	58-75	75-84

The Wall Street English Core Course was first launched in China in 2016 and has now been adopted by all of Wall Street English’s more than 420 learning centres worldwide.



# 4 / Purpose of the present project

The way in which the Core Course is constructed ensures that the instructional input received by learners is aligned to the GSE and CEFR. In other words, learners are following a syllabus which has been constructed based on a model of English language acquisition which has been empirically verified as accurate (North 2000; de Jong et al 2016). However, this fact does not guarantee that learners are acquiring the language capabilities in question, merely that they are presented with them.

There are two aspects to verifying language acquisition:

1. Programme Effectiveness: does studying at Wall Street English make a significant difference to learners' proficiency in English?
2. Accuracy of Language Acquisition Model: do the theoretically predicted gains in proficiency correspond to what happens in reality? (To what extent do the models of language proficiency at the various Wall Street English levels correspond to the reality of student ability at those levels?)

There are several reasons why it is important to establish answers to these questions:

- To give Wall Street English learners confidence that they are following an instructional programme which is effective and efficient.
- To be able to promise learners that they should be able to reach the objectives they have set themselves, and which have been promised to them in their contracts.
- To give Wall Street English staff confidence in the quality and accuracy of the instructional and assessment materials which they are administering.
- To ensure that inconsistencies between predicted and observed student performance can be used to improve and fine-tune the efficacy of Wall Street English instructional content and methods.

Wall Street English wanted therefore to be able to map student performance in as much detail as possible: eventually mapping and analysing student's actual performance (1) against their performance as predicted from their Wall Street English level, and (2) against individual learning objectives. The intention ultimately is to identify any areas of the curriculum in which learners underperform, and to modify the content or add remedial content accordingly. The present study is designed only to achieve the first aim (mapping actual against predicted student performance).

# 5 / Design and methodology

The study was aimed at answering the following primary questions:

1. How does the actual performance of Wall Street English learners compare with the target alignment of Wall Street English levels to the GSE/CEFR?
2. If there is a divergence, how does it vary by Wall Street English level?

Secondary questions to be answered in a future study were:

3. Can patterns be discerned in Wall Street English learners' strengths and weaknesses in performance across enabling skills (i.e. Reading, Writing, Speaking, Listening, Vocabulary, and Grammar)?
4. Are there patterns of strengths and weaknesses in learner performance across specific learning objectives?

The overall procedure was to test many Wall Street English learners of known levels using a test which would output a GSE score. This could then be plotted against the 'input' (Wall Street English instructional content of known GSE/CEFR level) to observe gains in language ability and to compare these with target or predicted values. Two criteria needed to be met for these results to be meaningful:

- It was necessary to ensure that learners had completed a minimum of two levels at Wall Street English - i.e. that instruction rather than placement was being assessed.
- The test needed to be one which not only reported GSE and CEFR levels, but also scores for enabling skills and for individual learning objectives.

The choice naturally fell upon Progress, as the only available test meeting both criteria. Moreover, as Wall Street English was then within the Pearson group, Progress could easily be made available, together with support in analysing the results. One additional factor was considered in designing the study: approximately half of students globally were in China, where the test was delivered from a different server; moreover, the size of many Chinese Wall Street English centres makes it relatively easy to carry out studies of this sort. The learners in the study were therefore placed into two groups: China and Rest of the World, with a view eventually to comparing overall and skill-level performance between the two groups.

# 6 / Procedure and data cleaning

## 6.1 Selecting the students for the study

A total of 107 centres were selected for the study, as follows: China 53, Czech Republic 1, France 15, Israel 8, Italy 12, Myanmar 1, Portugal 6, Saudi Arabia 3, Switzerland 1, Thailand 4 and Vietnam 3. To be eligible for the study, students had to have completed at least two Wall Street English levels since enrolling, so the lowest Wall Street English level at which students were eligible for the study was 3 (for students who had enrolled as zero beginners, at Wall Street English Level 1).

A total of 4252 students in these centres were eligible for the study and willing to take part: 1969 in China and 2283 in the rest of the world). Of these 109 either were duplicates (the same student taking the test twice) or had missing Wall Street English data (last unit completed and/or current level), leaving a total of 4143 students eligible for the test and taking it.

## 6.2 Taking the test and recording the results

Since Wall Street English students work at their own pace, and complete levels when they choose to, staff at the centres were asked to inform students when they became eligible to take a Progress test. Some students were already eligible when the study started because they had already completed at least two levels since enrolling at Wall Street English. Students did not necessarily take the Progress test immediately after completing a level, but when it was convenient for them and for the centre staff – so the last unit completed by each student was also recorded. This could have been 0, 1, 2 or 3 (usually 0 or 1) since the completion of the previous Wall Street English level.

An appropriate level of Progress was selected for each students' Wall Street English level, as follows:

Table 2: Progress test levels and Wall Street English levels

Progress version	15-30	25-40	35-50	45-60	55-70	65-80
Levels	3-4	5-7	8-10	11-13	14-16	17-20

The study was carried out over a period of 8 weeks in November and December 2017, during which time overall Progress scores and enabling skill scores (in Reading, Writing, Speaking, Listening, Vocabulary, and Grammar) were recorded for each student in question.

### 6.3 Cleaning the data

Many students who took the test (2528 out of 4143) had to be removed from the analysis because they had no overall Progress score, i.e. one or another of the 'low confidence' triggers was activated. As explained above, this could be for one of several reasons, the three major being:

1. The student started a test but did not complete it or did not complete sufficient items to receive a score. To remind the reader, the test was optional, and not a compulsory part of the Wall Street English curriculum.
2. The student was given a form of the test that was too difficult or too easy for them. As explained, if a score is below the 'floor' or above the 'ceiling' of the test in question, Progress does not return a numeric value.
3. The student's response was rejected for technology-related reasons, such as poor-quality sound on the voice recording exercise.

A summary of rejected test entries and reasons for rejection follows (data taken from the combined non-China or Rest of the World students):

Table 5: Reasons for rejected Progress test entries

Reason for rejection	Proportion of entries
Low confidence + Score below range + Insufficient items in section 1	12.2%
Score below range + Insufficient items in section 1	28.4%
Insufficient items in section 1	51.0%
Low confidence + Score below range	1.2%
Score above range	1.0%
High overall error of measurement	0.1%
Other (low confidence, insufficient items, theta calculation error, score below range)	6.1%

This set of 1615 student records was then analysed as described in the next section, to verify the extent to which learners' scores corresponded to what would be expected from their Wall Street English level.

An additional refinement was then carried out to corroborate the previous results. Most students (81%) scored zero in one or more six enabling skills (Vocabulary, Grammar etc.), because they failed to complete that part of the test or for some other technical reason:

Table 6: Breakdown of missing subskill scores

Missing subskill scores	Number of records	Percentage of total
All subskill scores missing	10	0.6%
5 subskill scores missing	60	3.7%
4 subskill scores missing	317	19.6%
3 subskill scores missing	335	20.7%
2 subskill scores missing	340	21%
1 subskill score missing	293	18.1%
All subskill scores present	260	16.1%

A new data set was created, composed only of the 260 student records where all six subskill scores were present. The hypothesis was that this smaller set should have yielded more accurate results, as the students' Progress scores represented the most complete assessment possible of their performance in the test. These students' scores are also least likely to have been negatively affected by low motivation. The analysis described in the next section was accordingly carried out on the reduced data set too, as a consistency check on the previous analysis.

To summarise the process by which the data sets for analysis were created:

Table 7: Summary of process of creation of data sets

Item	Total	Percentage
<b>A.</b> Total students eligible for test and taking it	4143	100.0%
<b>B.</b> Removed because lacking overall Progress score	2528	61.0%
<b>C. Data set 1 for analysis</b> (= A-B)	<b>1615</b>	<b>39.0%</b>
<b>D.</b> Removed because lacking one or more Progress sub-skill scores	1355	32.7%
<b>E. Data set 2 for analysis</b> (= C-D) (all subskill scores present)	<b>260</b>	<b>6.3%</b>

# 7 / Analysis and results

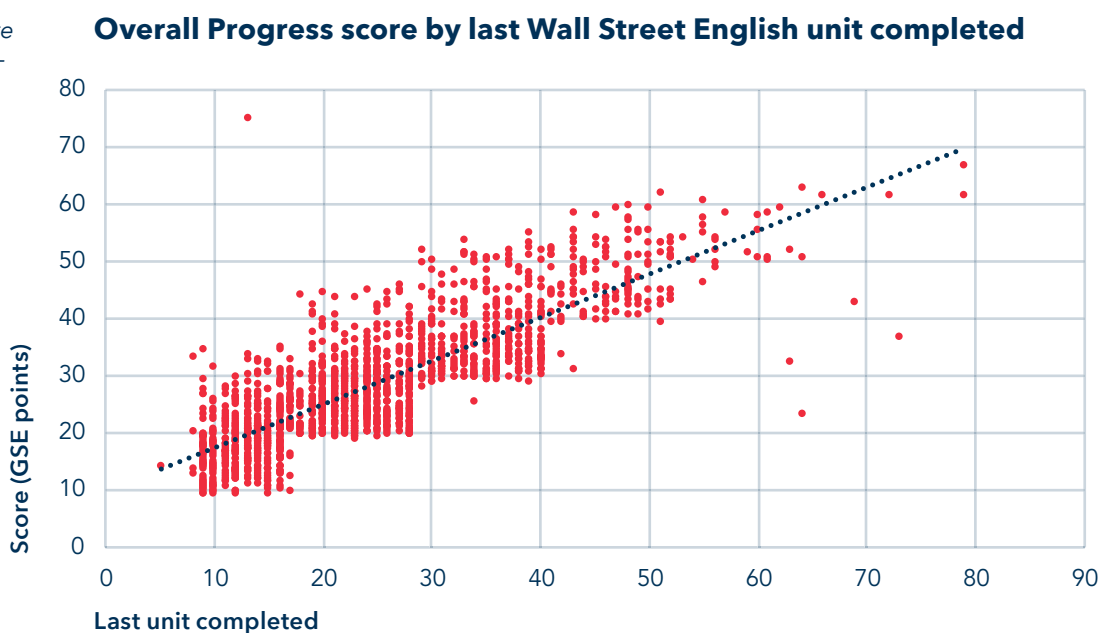
## 7.1 Overview

The method chosen for evaluating the data was regression analysis. This is a technique which looks at the relationship between a dependent variable and one or more independent variables, to determine to what extent the value of the dependent variable can be predicted from that of the independent variable(s). The aim of the analysis is to determine how much of the variation in the dependent variable is explained by variation in the independent variable(s).

## 7.2 Analysis of data set 1

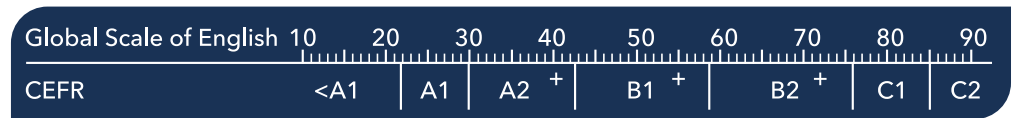
In the analysis of the first data set (all students with complete Wall Street English records and non-zero overall scores in Progress), students' Progress scores were plotted against their last completed Wall Street English unit, rather than their current level (i.e. last unit out of 80, rather than current level out of 20). The hypothesis was that this would model the distribution of student performance levels more accurately. The resulting scatter chart is shown below, including the trendline:

Figure 2: Progress score by last unit completed - scatter chart



To remind the reader of the relationship between GSE scores and CEFR levels:

Figure 3: Global Scale of English and CEFR



It will be noted that:

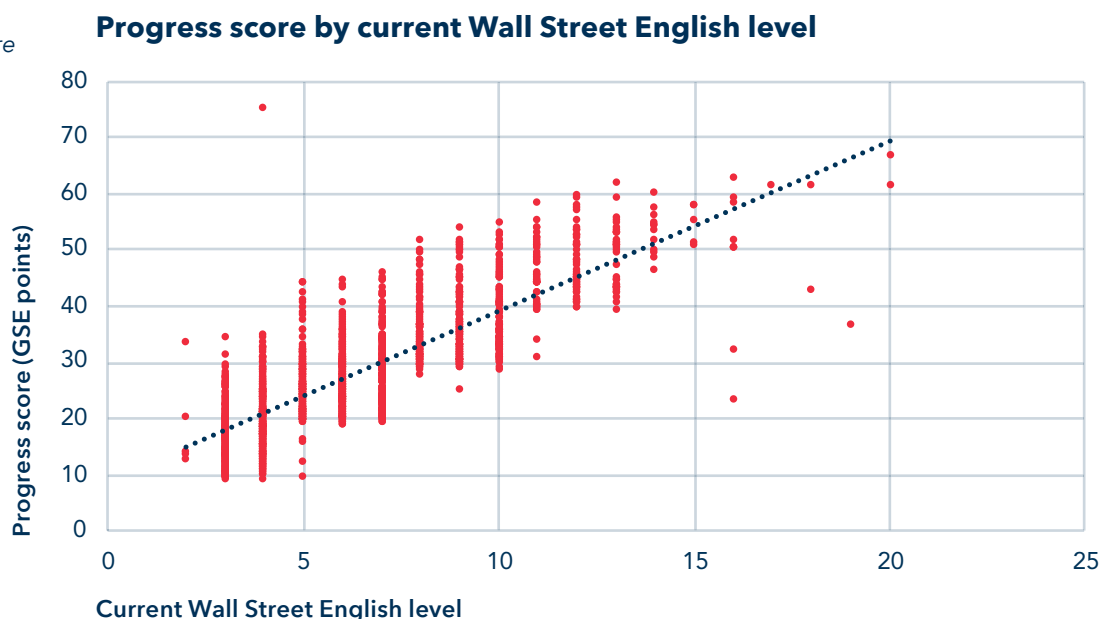
1. there is a large concentration of students at lower levels (Units 9-40) and very few indeed at levels above 14 (unit >55), and
2. there is a 'step effect' at Progress scores 20, 30 and 40. This is the result of using a different version of Progress with students at these levels, with a different 'floor' at each level.

As explained above, students scoring below the 'floor' have no score assigned to them, hence are simply removed from the data set.

The coefficient of determination between the two variables ( $r^2$ ) was found to be 0.6861, on a scale from 0.0 (no relationship) to 1.0 (value of the dependent variable can be fully predicted). Given the sparseness of the data at the higher levels, with the corresponding risk that the results would be distorted by sampling error, the analysis was re-run on students at Units 1-40 (Levels 1-10) only. This resulted in only a very modest and insignificant increase in  $r^2$ . As the chart in Figure 2 shows, the spread of data is relatively high at all units/levels.

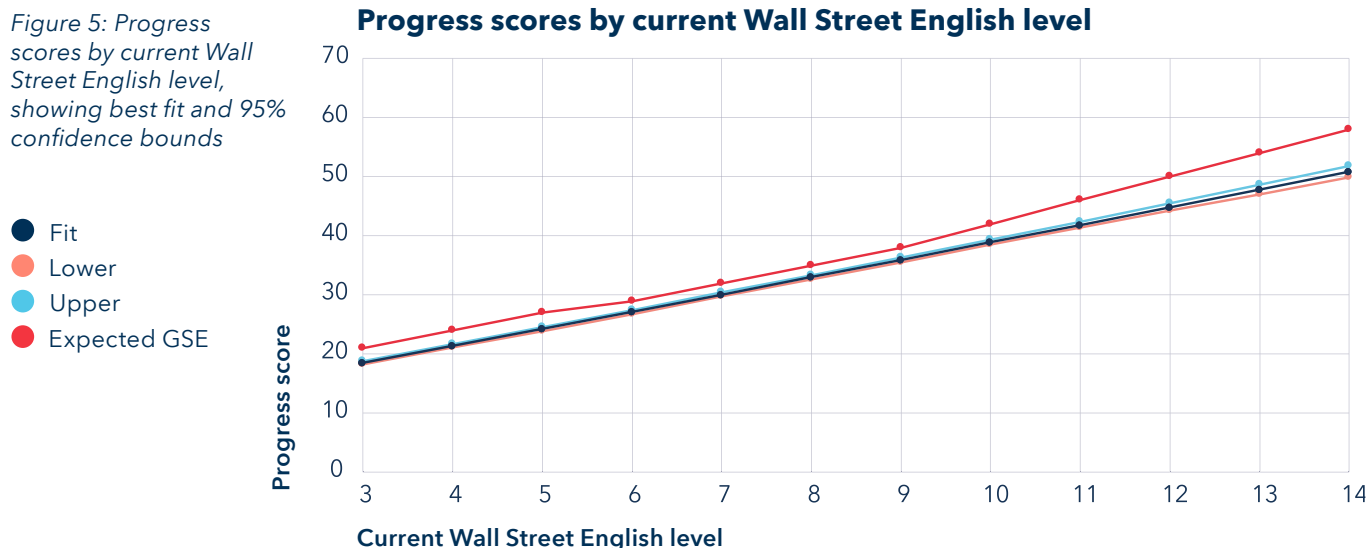
Since the error of measurement of Progress ( $\pm 3$  GSE points) is considerably greater than the theoretically established difference in GSE between one Wall Street English unit and the next (1-1.5 GSE points), the difference in Progress score across adjacent Wall Street English units may be too greatly affected by measurement error. It was hypothesised that this may have caused the high degree of variability in the results. Thus, the team decided to carry out a further analysis, plotting students' current Wall Street English level against their Progress score. The results show a similar clear trend and degree of correlation:

Figure 4: Progress score by Wall Street English level



The coefficient of determination between the two variables ( $r^2$ ) was found to be considerably higher, at 0.92, after fitting a model predicting Progress score from Wall Street English level, with linear and quadratic terms for Wall Street English level. The following chart shows the line of best fit, with upper and lower bounds of 95% confidence (i.e. the probability that a confidence interval constructed this way covers the correct value is 0.95):

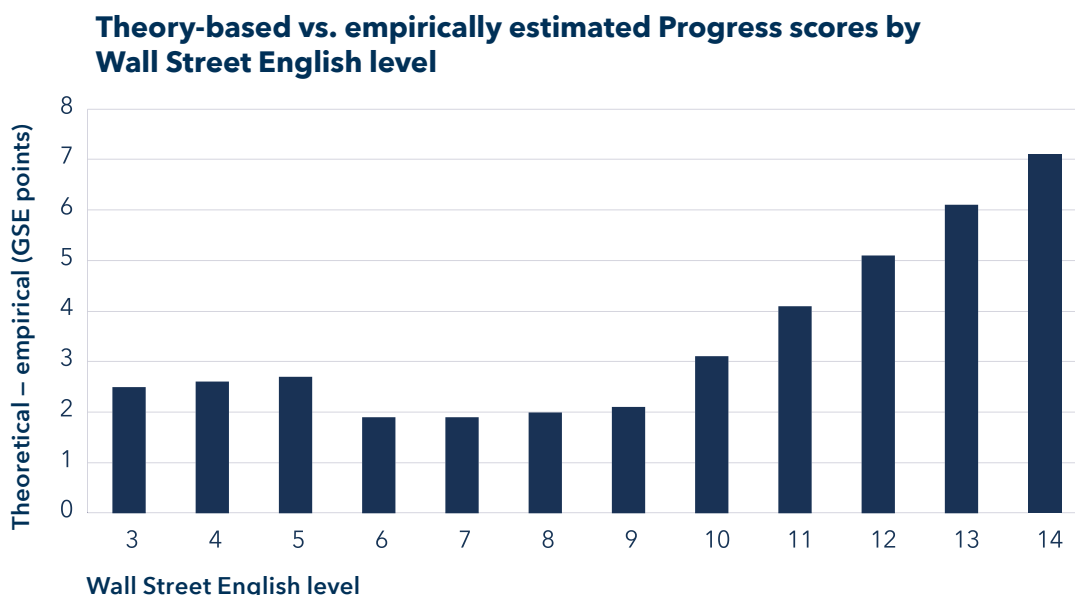
Figure 5: Progress scores by current Wall Street English level, showing best fit and 95% confidence bounds



This chart shows that the line is a very good fit to the data, with upper and lower bounds extremely close to the fit all the way up to Level 14 (there was insufficient data to go beyond this point). This shows that there is a consistent and regular increase in students' mean Progress scores as one goes up the Wall Street English levels.

However, it will be noted that the line of fit is consistently below the theoretically-established GSE value for students at all Wall Street English levels. The students' actual performance falls consistently below the published alignment of Wall Street English levels to CEFR/GSE, as predicted from the Learning Objectives on which the course design is based (as described in section 3 above). Here is a breakdown by Wall Street English level:

Figure 6: Difference between theoretically established and observed Progress scores





Although the difference is below the error of measurement of Progress ( $\pm 3$  GSE points) up to and including Level 10, it is still significant because the differences are not randomly clustered around zero but are consistently positive. In addition, these are averages across numbers of students, which should be less affected by measurement error.

The team hypothesised that incomplete data might have been responsible for this: as explained in section 6.3 above, over 80% of the overall Progress scores in data set 1 were based on incomplete evaluations, because they were missing a score in one or more subskills. There are many reasons why a score may not be recorded, as explained in section 6.3, but it is likely or probable that:

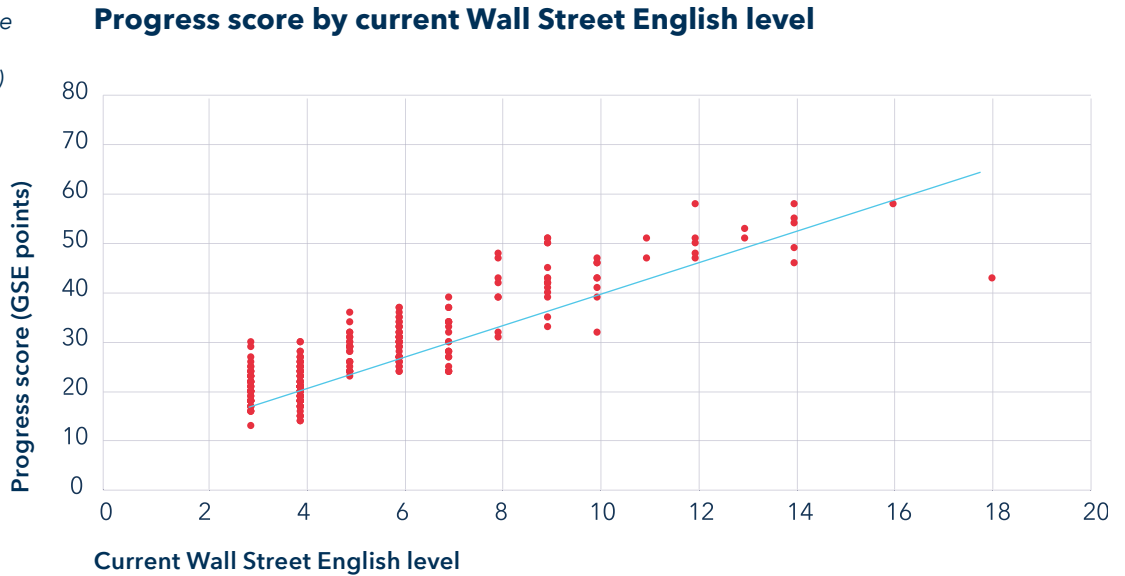
- a. the group of students who did not complete the test includes a higher proportion with below average ability than those who did complete it, and/or
- b. the error of measurement for students who missed or failed to complete one or more sections of the test is higher than for those who did complete it.

Therefore, the team decided to run a second analysis on the much smaller set of student data (260 records) with scores in all subskills, to ascertain whether the line to data fit would be similar to that for the full set, and particularly if the difference between the theoretically-established and empirically-estimated Progress means would be similar, or less, or greater.

### 7.3 Analysis of data set 2

The team plotted Progress score against current Wall Street English level, as in Figure 3 above. This resulted in the following scatter chart, starting as usual from Wall Street English Level 3:

Figure 7: Progress score by Wall Street English level (reduced data set)

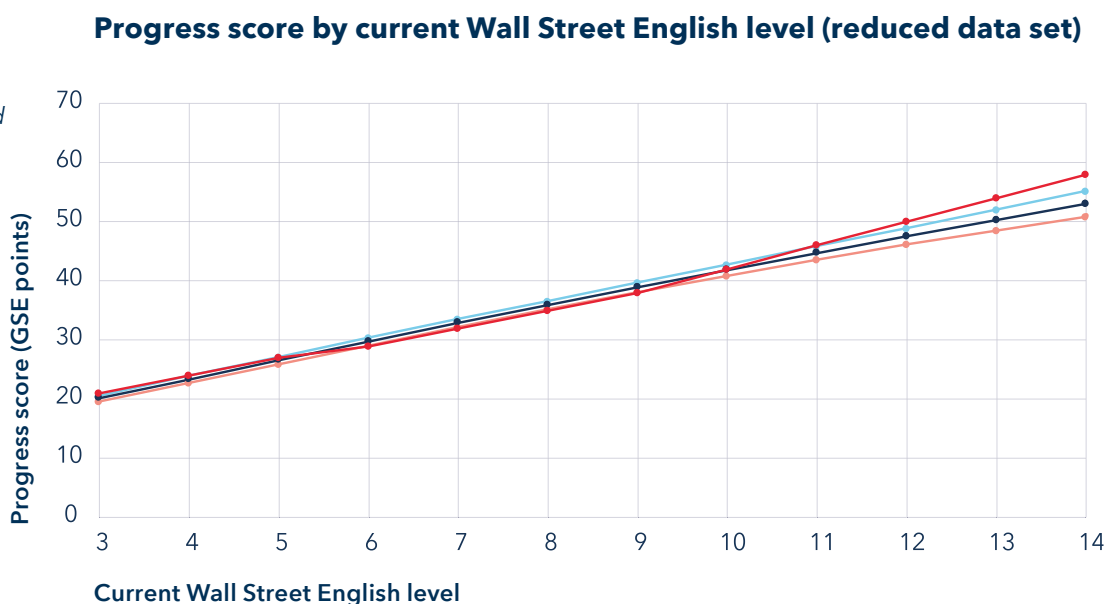


The same step effect is present, for the same reason (change of Progress test), and the sparseness of data at the higher levels is even more evident. The correlation of the two variables is slightly closer than in the larger dataset:  $r^2$  (the coefficient of determination) for this analysis was 0.95, after fitting the same regression model as before with linear and quadratic Wall Street English terms.

The following chart shows the line of best fit, with theory-based ("expected") and empirically-estimated ("fit") values for Progress scores by Wall Street English level up to level 14, with upper and lower bounds of 95% confidence (as stated, there is insufficient data to draw firm conclusions about higher levels at this point).

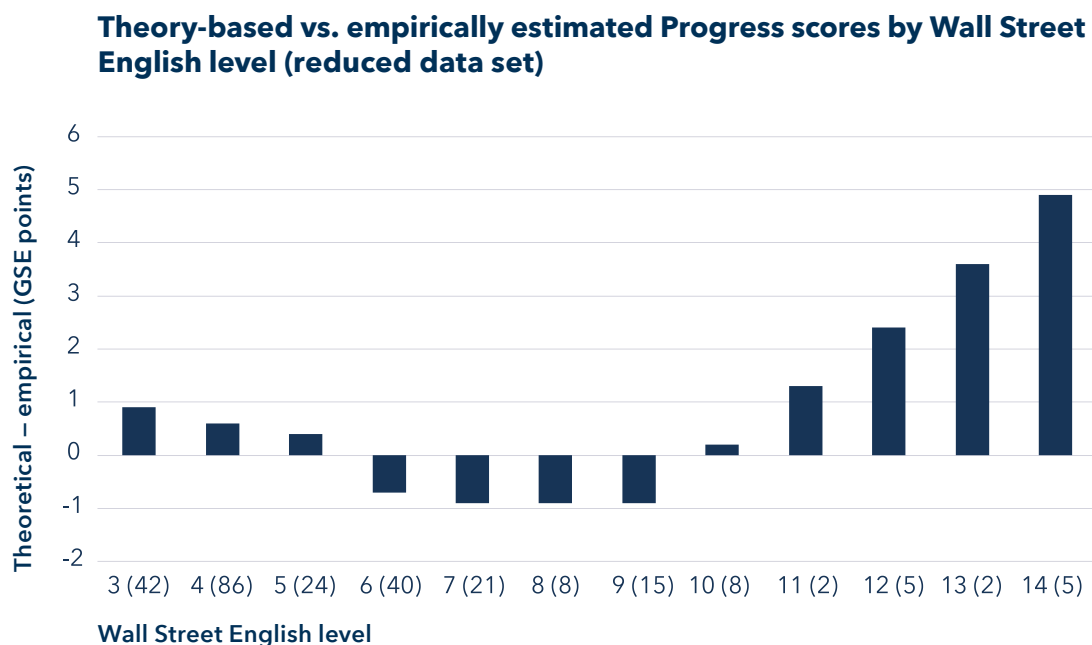
Figure 8: Progress score by Wall Street English level – best fit and expected (reduced data set)

- Fit
- Lower
- Upper
- Expected GSE



If this chart is compared with Figure 5, it will be noted that predicted and observed values, together with upper and lower bounds, are very close together at the lower levels, though they start to diverge above Level 12. Here is a chart of the differences between the theory-based and empirically-estimated values by Wall Street English level, for comparison with Figure 5, showing also the number of students at each level:

Figure 9: Difference between predicted and observed Progress scores (reduced data set)



The differences are now clustered around zero, and only rise above the error of measurement of Progress ( $\pm 3$  GSE points) above Level 10 - where it will also be noted that there is a very small number of student records.

## 8 / Conclusions and next steps

The combination of the two analyses (extended and reduced data sets) suggest firstly a significant and consistent increase in Progress scores going up the Wall Street English levels, and secondly a good fit between observed and predicted scores for Wall Street English students, corroborating the published GSE/CEFR equivalents of Wall Street English levels as an accurate reflection of student performance up to and including Level 12.

Regarding the difference between the two analyses, the team's hypothesis (that the apparent consistent underperformance of Wall Street English students compared with their predicted scores was related to missing Progress subskill scores in the larger data set) seems to be verified by the absence of this phenomenon in the reduced data set. When all student records with incomplete Progress scores were removed from the data set for the purposes of the second analysis, the weighted average difference between theoretically established and empirically estimated scores from Level 3 through Level 12 was 0.44.

When looking at the higher levels (13-20) there appears to be the beginning of a significant divergence between expected and fit scores, which is more marked as one goes up the levels from 14: a number of extreme outliers can be seen in Figure 1. However, there is very little relevant data even in the larger data set: 139 students in Levels 10-14 but only 20 students in levels 15-20. Of these, 76% have two or more missing subskill scores and are therefore of questionable validity. In the reduced data set (students with Progress scores in all subskills) there were 17 students in Levels 10-14 and only 2 in levels 15-20.

Note also that, since Core Course has been in use for under two years, students at higher levels have not followed it from the start. Those who enrolled at Wall Street English at a much lower level will have followed the previous Wall Street English course, while those who enrolled at a level close to their current level will have done much of their study elsewhere. In either case the performance of these higher-level students in Progress may not be as accurate a reflection of the effectiveness of Wall Street English's current instructional programme (i.e. Core Course) as we would like. It would be useful to carry out a further study, perhaps using a different testing instrument, to assess the accuracy of the alignments to GSE/CEFR of Wall Street English Levels 15-20. Note, however, that fewer than 4% of current Wall Street English students are in these levels, so the alignment of the lower levels is of much more practical consequence.

# References

Buckland, S. Using the CEFR to benchmark learning outcomes: a case study. EALTA, 2010.

Available at [www.ealta.eu.org/events/Barcelona2010/Buckland.pdf](http://www.ealta.eu.org/events/Barcelona2010/Buckland.pdf)

Council of Europe (2001) Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: CUP.

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 <https://online.flippingbook.com/view/894261/>

North, B. (2000) The development of a common framework scale of language proficiency. New York: Peter Lang.

Pearson Inc. (2017) PTE Academic.

Available at <https://pearsonpte.com/the-test/>

Pearson ELT (2018a) Global Scale of English.

Available at <https://www.pearsonelt.com/about/gse.html>

Pearson ELT (2018b) Progress.

Available at <http://product.pearsonelt.com/progress/#.WsS9RojwZaQ>

Pearson ELT (2015) Progress Technical Report. Available at

<https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/SupportingDocs/Progress-test-information.pdf>

van Ek, J. and Trim, J (1979 & 1990) Threshold. Strasbourg, Council of Europe, and Cambridge University Press. Purchasable at

<https://book.coe.int/en/root/805-threshold-level-1990.html>



**For inquiries: Bindi Clements**  
Instructional Design and Efficacy Manager

**[bindi.clements@wallstreetenglish.com](mailto:bindi.clements@wallstreetenglish.com)**