Pearson

Research Series

Developing GSE Grammar

Intermediate Report

May 2017

Simon Buckland
Prof. John de Jong

Global Scale of English
Fast-track your progress

Contents

Executive summary

Global Scale of English (GSE) Grammar is designed to provide detailed information for teachers on the key enabling skill of grammar, linking grammatical structures to their usage pattern, related language functions, and GSE/ Common European Framework (CEFR) (Council of Europe, 2001) levels. This newly created GSE-linked inventory of grammar structures aims to bring together a range of information not currently available in existing grammar syllabuses or textbooks:

- Structures are given precise levels (GSE values) as well as CEFR levels.
- Structures have been levelled empirically.
- Illustrative examples are provided are for each grammatical structure.
- Related structures are grouped together for comparison and also listed separately.
- Grammatical structures are linked to related functional learning objectives.

GSE Grammar was created from a range of authoritative sources, including the Council of Europe's four English language syllabuses. Can Do descriptors were written for each grammatical structure, to ensure conformity with the GSE Learning Objectives for Adult Learners. These descriptors, with their illustrative examples, were then rated for usefulness by experts, in workshops and online, and the resulting ratings analysed statistically to compute the ranking and GSE values for all the descriptors.

GSE Grammar can be found on english.com/gse as part of the GSE Teacher Toolkit, and can be downloaded at www.english.com/blog/download-gse-grammar.

Purpose of the project

This report describes the development of GSE Grammar. GSE Grammar is an inventory of grammar structures designed to sit alongside the GSE Learning Objectives for Adult Learners, providing detailed information on the enabling grammar skills that are necessary for successful communication. It links grammatical structures to their usage patterns, related language functions, and GSE/CEFR levels.

GSE Grammar is designed to enable teachers to understand at which level grammar structures or groups of structures are necessary for successful communication in English. It also helps them to understand and define the usage of an unfamiliar or partially understood grammar structure, by providing examples of the structures in use.

This report assumes a basic understanding of statistics and testing terminology, although a glossary is included for reference.

The development of GSE Grammar is an ongoing process, and the learning objectives described in this document will be extended in future updates. GSE Grammar is to be seen as an inventory of structures ranked by usefulness for learners, rather than an instructional sequencing based on ease of acquisition. In future stages of the project there will be acquisition-based validation of structures' assigned level of difficulty, carried out both by Pearson staff and independent researchers and teachers.

For more information about our research programme, please visit english.com/gse/researchers.

URL DOES NOT WORK

Background

The Global Scale of English (GSE) is a standardised granular scale which measures English language proficiency. It is designed to support a more fine-grained understanding of proficiency than is possible with other frameworks such as the Common European Framework of Reference (CEFR). The GSE has been psychometrically aligned to the CEFR.

GSE Grammar is a subsidiary element within the Global Scale of English ecosystem. This ecosystem is 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 describe what it means to be at a level of proficiency in English in terms of (a) what a learner can do at that level – these are the Functional Descriptors – and (b) the enabling skills of Vocabulary and Grammar required to perform a particular language function. In line with the model used for the levelling of functional GSE Learning Objectives, a learner is considered to be at a given level is they can perform 50% of the tasks that characterise that level.

Four sets of GSE Learning Objectives have been developed, each tailored to meet the needs of specific audiences: Adult Learners of General English, Learners of Academic English, Learners of Professional English and Young Learners (6–14). GSE Grammar has been developed with adult learners in mind, suitable for those studying general, professional and academic English. Full information about the Global Scale of English is available on our website: english.com/gse.

Part 1 Creating GSE Grammar

Grammatical structures were extracted from a variety of authoritative sources including the Council of Europe's language syllabuses and Pearson's own courseware to create an initial inventory of 437 grammar learning objectives.

GSE Grammar is intended to be a pedagogical grammar aimed at teachers of English, rather than a reference or descriptive grammar aimed at linguists or other academic specialists. This distinction is important and has implications for the components and composition of GSE Grammar. As a result, the following decisions were taken:

- to use terminology with which teachers (and learners) are familiar rather than technical grammatical terms (e.g. 'past simple' rather than 'preterite')
- to include the same grammatical forms and to classify them in a similar way as the grammars and coursebooks with which teachers and learners are familiar
- to use teacher judgements to determine the GSE levels of grammatical structures
- to describe grammar structures in terms of their communicative effect as well as their form

These last two decisions are consistent with the approach taken in the development of the functional GSE Learning Objectives which are presented in the form of Can Do statements and were rated for proficiency level by teachers and pedagogical experts.

Grammatical forms were extracted from a range of authoritative sources:

1. The Council of Europe's language syllabuses: *Breakthrough* (A1), *Waystage* (A2), *Threshold* (B1), and *Vantage* (B2). These contain tables of Functions, General Notions and Specific Notions as well as grammatical exponents characteristic of Functions and General Notions only.
2. Pearson's own student grammars: MyGrammarLab and Focus on Grammar – used to identify possible structures not found in the Council of Europe's syllabuses, especially higher-level forms.

These sources provided provisional CEFR levels for each structure, according to (1) the Council of Europe syllabus or (2) to the Pearson textbook in which they first occur.

Each structure in the GSE Grammar inventory contains the following information:

- a Learning Objective (Can Do statement), in the format: *Can use structure X to perform task Y*
- a 'label' corresponding to the term(s) normally used in grammar and course books to refer to the structure
- examples of the structure in use

All grammar learning objectives were reviewed for consistency and clarity by in-house editors before being passed to teachers for rating.

Part 2 The rating process

Teachers and pedagogical experts trained in a series of workshops were asked to assess structures on a five-point scale of usefulness. The structures were also rated by a group of over 900 online raters.

Grammatical structures were rated on a 5–point scale for usefulness to learners, rather than by CEFR or GSE levels. This was to encourage the experts to think 'out of the box' rather than reproduce the conventional levels usually found in textbooks. The custom scale is shown below, in the form in which it was presented to raters.

A team of 55 experts were trained in a series of webinars, using six representative structures from the inventory, two of which were rated collectively in group discussion and the remaining four individually and then in discussion. The results of these training sessions were very encouraging: only in one case was there a spread of more than 1 point on the 5–point scale among the ratings given in the workshops.

The 437 structures were divided into four batches, with ten structures (anchor items) shared among all batches, so that each rater was given approximately 120 structures to rate per batch. Some raters rated more than one batch.

A further 952 raters (mostly experienced teachers) were recruited via extended contact networks, and were asked to rate online via Survey Monkey. This group was not formally trained, and were given only the rating scale. The online survey ran for three weeks. The four batches of structures were all sub-divided into 6 overlapping sets, so that each of the 'online raters' was given approximately 40 structures to rate.

Number	Label	Meaning	Examples
1	Fundamental	I'd teach this right at the beginning of a course	Verb '*be*', singular, present tense: *am, is, are*
2	Essential	To be taught early on, but not right at the beginning	Object pronouns: *me, him, her, us, them*
3	Useful	Everyone needs to learn this, but it isn't among the first things to learn	Present perfect of verbs
4	Nice to have	Good to learn this as part of an extended course	'Parallel' comparatives: *the more she earns, the more she spends*
5	Refinement	A very advanced or 'nuance' structure which could be left out altogether	Inversion after only + conjunction: *Only when the temperature rises above 50 does the plant flower.*

Part 3 Analysing the ratings

The ratings were analysed statistically and combined with the CEFR levels assigned by the Council of Europe to produce a ranking of structures from the most to least useful for learners. These rankings were then mapped to the Global Scale of English to determine GSE values for each structure.

Data cleaning

For each descriptor, the average rating per rater group (experts and online raters) was computed separately. An unreliable rating (an error of judgement or a slip of the hand) was deemed to be one which deviated more than 1.5 from the average for that descriptor within that group (expert or online). Thus the first stage of cleaning consisted in removing all such deviating ratings from the data set. Approximately 3% of expert ratings and 12% of online ratings were removed for lack of agreement.

Certainty of ratings

For the remaining ratings, a certainty index was computed for each descriptor: the proportion of ratings for that descriptor which fell within two adjacent categories. The average certainty index for the expert ratings was 0.90 and for the online ratings 0.81.

Experts and online raters agreed moderately well ($r^2 = 0.68$) on all 437 descriptors. When all flagged (possibly problematic) descriptors were removed, the remaining 229 descriptors with zero flags raised r^2 to 0.84 (see below for further explanation).

Combining Expert and Online ratings

Ratings from experts and online raters were combined by weighting them according to their certainty as follows:

$$\text{Combined rating} = \frac{R_E.C_E + R_O.C_O}{C_E + C_O}$$

where R_E is the average rating by experts, C_E is the average certainty of expert ratings, R_O is the average rating by online raters and C_O is the average certainty of online ratings.

Flagging ratings

Ratings were flagged for one of seven possible reasons:

1. There were too few expert ratings (<16 for a single descriptor).
2. There were too few online ratings (<10 for a single descriptor).
3. The certainty of the expert ratings was less than 0.7.
4. The certainty of the online ratings was less than 0.7.
5. The standardized difference (z-score) between expert and online ratings was below –1.96 or over +1.96.
6. No CEFR level drawn from one of the Council of Europe syllabuses (*Breakthrough*, *Waystage*, *Threshold*, or *Vantage*) was available.
7. The standardized difference (z-difference) between the average rating and the Council of Europe level was below –1.96 or over +1.96. (Note: this flag could not be set if flag 6 was positive).

The final published set excludes descriptors with more than one flag. The graph below shows how the fit of only unflagged descriptors is better than that of all descriptors (there are fewer outliers).

All ratings: Expert vs Online

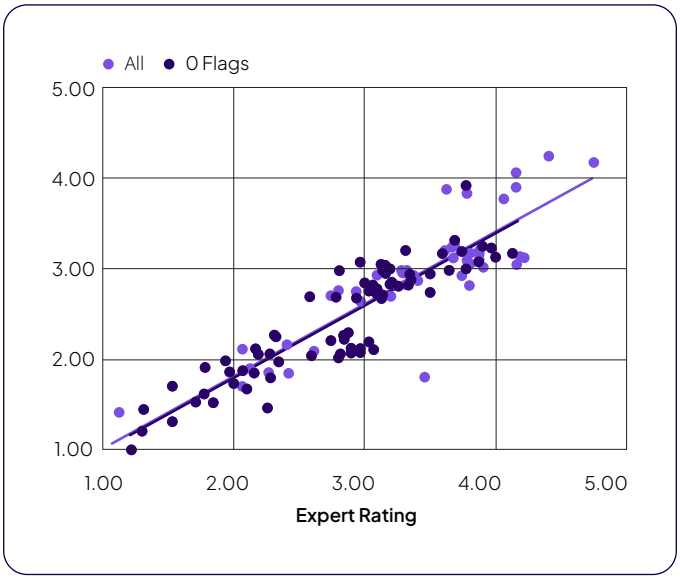

Figure 1: Comparing online and expert ratings for all descriptors with unflagged descriptors.

From ratings and Council of Europe levels to GSE values

A transformation function from descriptor combined-rating to the Council of Europe logit scale was then estimated, based on the 229 descriptors which had no flags.

The starting point was the Council of Europe level for each structure, according to which syllabus it first occurs in. Cut-off s for each CEFR level are given in North (2000), expressed in logits. These are therefore the lower boundaries of each CEFR level. Assuming that, for example, Threshold grammatical structures occur from the 'bottom' to the 'top' of Threshold, a fair estimate of the average Threshold structure would be at the midpoint between the lower and upper boundary of the Threshold interval (B1) on the CEFR scale.

A best-fitting regression function was then estimated between these CEFR midpoints and the combined value of the collected ratings, computed as described above. The best-fitting function (explained variance 0.68) was a second order polynomial:

$CEFR = 0.711 \times CRV_2 + 0.934 \times CRV - 5.0925$, where CRV is the combined rating value.

This transformation function was subsequently applied to all descriptors. The following table shows the distribution of non-flagged structures across levels, with the peak concentration in GSE 43–51 (B1) as one might expect. Grammatical knowledge plays an increasing role from level A1 onwards, and the number of grammar points needed for successful communication increases up to Level B1. After this the number of additional grammar points per level starts to decrease, and by level C1 grammar seems to be largely mastered.

GSE	CEFR	# Structures
22–29	A1	42
30–35	A2	44
36–42	A2+	69
43–50	B1	129
51–58	B1+	92
59–66	B2	42
67–75	B2+	17
76–84	C1	2
85–90	C2	0
	Total	**437**

Part 4 Validating the ratings

When rating the GSE Learning Objectives (functional descriptors) we were able to use descriptors from the CEFR already rated by North (2000) as anchors. We had no such anchors available to use in GSE Grammar.

However, connections between some of the grammar descriptors and specific GSE Learning Objectives can be identified, meaning that the grammar structure can be understood to be required to correctly perform the functional objective. For example:

Grammar

Can use 'can' to refer to ability in the present and near future. (GSE 29)

GSE Learning Objective

Can express ability or lack of ability with regard to basic activities using 'can' or 'can't'. (GSE 27)

Such connections could be identified for a total of 40 GSE Learning Objectives. These were used for validation of the ratings of their corresponding grammar structures.

The GSE values of the 40 GSE Learning Objectives which were chosen as anchors were compared with the ratings of the corresponding grammar structures. There were two misfits, but the remaining grammar descriptors were close to their corresponding learning objectives with an r2 of 0.83. On average, the grammar descriptors were extremely close to their corresponding functional descriptors: just over 2 GSE points higher or lower.

Only Acceptable pairs; one misfit removed

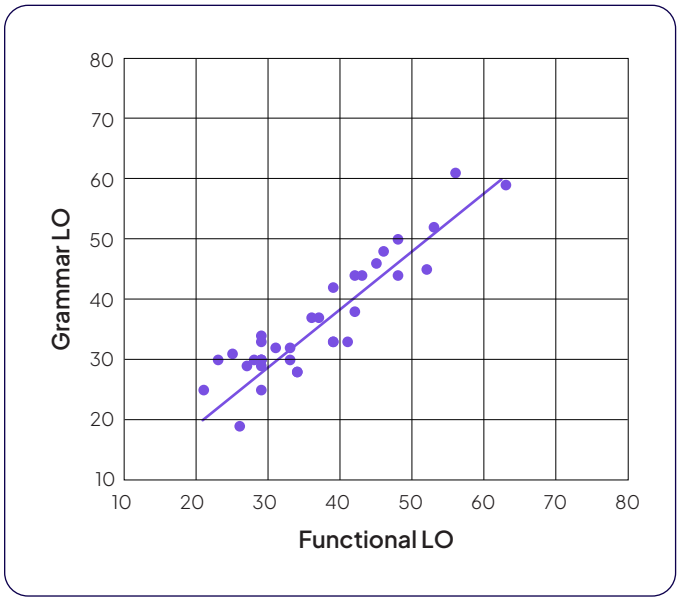

Figure 2: Comparing GSE values for grammar structures with corresponding GSE Learning Objectives

Part 5 Linking to GSE Learning Objectives

Surveys of potential users suggested an interest in accessing links between grammar structures and functional objectives. Where lesson or curriculum goals are defined in -terms of functional objectives, teachers and content developers need to determine what are the relevant enabling skills. This observation justified our decision to establish reciprocal links between grammar structures and functional objectives wherever possible.

Surveying the functional objectives from this perspective, we found that they could be grouped into three categories:

1 Quasi-grammatical in nature, dependent on structures named or clearly implied in the descriptor itself.

For example: *Can answer simple questions about the location of people or things in a limited way. Can compare quantities in a basic way.*

2 Performance descriptors, or those potentially realisable using too wide a range of language forms to be linkable to specific structures.

For example: *Can exchange simple information on everyday topics, provided the other person speaks slowly and clearly and is prepared to help.*

Can initiate, maintain and close simple, restricted face-to-face conversations

3 Descriptors potentially realisable with a limited number of specific structures, but which are not named or clearly implied in the descriptor itself.

For example: *Can discuss what to do in the evening or at the weekend.*

Can narrate a story.

Functional descriptors in categories 1 and 3 should be linkable to grammatical descriptors, but the above observation was not sufficient to generate the required links on an objective or evidential basis:

- The categorisation itself is inherently subjective, as the categories overlap with each other, and
- To select the matching grammar descriptors for functional descriptors in category 3 (probably the largest one) 'by hand' would also be an arbitrary and subjective process.

Given the non-systematic and highly contextual relationship between form and function in language, the connections which we required needed to be made indirectly, via intermediate links established on some sort of objective or expert-reviewed basis. Such links were fortunately available, in the shape of the Council of Europe Vantage Functions and (General) Notions.

As explained in the document itself (van Ek and Trim 2001, p.22), Functions are "*… the kind of things people may do by means of language*" and General Notions are "*… [those] concepts that we may refer to while fulfilling language functions … [which] may be expressed in almost any situation*".

The Vantage syllabus minutely classifies Functions and Notions, and also links them to grammatical exponents.

For example:

A2 **Pronouns**

2.1 **Types of pronoun**

2.1.1 demonstrative (5.1.1, 6.8.1.1–2) **this, that, these, those**

2.1.2 personal (5.1.1, 6.8.1.1–2)

2.1.2.1 subject forms **I, you, he, she, it, we, they**

2.1.2.2 non-subject forms **me, you, him, her, it, us, them**

2.1.3 possessive (6.8.1.1–2) **mine, yours, his, hers, ours, theirs** (ibid., p. 144)

The numbers in brackets are references to Functions (5.x) or General Notions (6.x), and these references were imported into the database for the GSE grammar learning objectives.

There are, however, no established links between functional learning objectives and Functions and General Notions – not even for the 30% or so of the former which were taken verbatim from the CEFR. However, a prima facie connection is generally easy to establish: provided the language function or the topic (notion) specified in the descriptor is sufficiently specific, it is possible to identify the Function(s) or General Notion(s) needed to realise the task described.

For example: *Can compare and evaluate ideas in a structured and logical text.*

was determined to be related to two Vantage functions:

1.5.2.5 **giving information: reason,** and

5.22 **introducing a counter-argument,** and one General Notion:

4.3.1 **degree: comparative forms**

In other words, in order to *compare and evaluate* ideas, it is necessary to *give information (reasons)*, to *introduce counterarguments*, and to *make comparisons*.

In conclusion, each grammar descriptor has been linked to one or more Functions or General Notions, and each functional learning objective in categories 1 or 3 as described above has also been linked to one or more Functions or General Notions. These indirect links are used to determine which grammar descriptors enable a particular learning objective, or vice versa.

The indirect links underlie the links that are made visible but produce an additional level of complexity. For this reason they have not been made visible to users.

To give a further example, take the grammar descriptor *Can use 'one of/some of/among' in phrases with superlative adjectives. (one of the best schools; among the richest people; some of the finest examples).*

A user searching for related learning objectives (general adult) will find the following:

Can compare and evaluate different ideas using a range of linguistic devices.

Can compare the advantages and disadvantages of possible approaches and solutions to an issue or problem.

Can use a range of language to make detailed comparisons of quantities.

The intermediate Function and General Notion links are not made visible to the user, as they introduce an unnecessary level of complexity.

Part 6 Discussion: future developments

The degree of certainty, both within the ratings of the structures and between the ratings and external measures such as the Council of Europe levels and the ratings of related GSE Learning Objectives for Adult Learners, suggests that the GSE values attributed to the structures are reliable.

Further validation will be carried out as part of our ongoing research programme by looking at the results of the grammar component of online tests such as Progress and comparing learners' actual ability to produce structures with their expected performance based on the GSE levels in the inventory.

GSE Grammar is available to download at english.com/gse. It forms part of the GSE Teacher Toolkit – see english.com/gse/teacher-toolkit.

Glossary

Term	Definition
CEFR	Common European Framework of Reference for Languages.
certainty value	The proportion of ratings within two adjacent categories on a categorical scale.
correlation	A statistic showing the interdependence between two variables.
descriptive (reference) grammar	One which describes the grammar of the language in exhaustive detail, from the perspective of an academic specialist in language.
explained variance	The explained variance in the context of a regression analysis is the fraction of the variance of the dependent variable which can be explained by the variance of the independent variable.
general notion	Abstract or conceptual topics and topic areas, such as Distance and Sequence, which are likely to be encountered in any situation. (Compare specific notion).
GSE Learning Objective	A description of what a student is likely to be able to do at a particular point on the Global Scale of English.
logit	A logarithmic function used in the logistic model, which is commonly used to estimate the probability of a binary response based on one or more independent variables.
pedagogical grammar	A grammar which describes how to use the grammar of a language to communicate, for learners and teachers of the target language.
r^2 (coefficient of determination)	The symbol for explained variance (see above). r^2 is a measure of how well a set of data fits a statistical model – usually a line or a curve. An r^2 of 1 indicates that the data perfectly fit the line or curve, while an r^2 of 0 indicates that the data does not fit the model at all. Anything above 0.8 is generally regarded as an acceptable fit.

regression function	A mathematical function expressing the relation between a dependent variable (y-axis) and an independent variable (x-axis).
specific notion	Concrete and specific topics and topic areas, such as Food and Drink and Clothing, which are encountered only in specific situations. (Compare general notion).
standard deviation (SD)	A statistic showing the amount of variation in a data set. A standard deviation close to 0 means all data points are close to the mean.
transformation function	A mathematical function allowing values on one scale to be transformed to corresponding values on another scale.
z-score	A z-score or standardized score is a statistical measurement of a score's relationship to the mean in a group of scores expressed in standard deviations of these scores to the mean. For example, a z-score of 1 is one standard deviation above the mean and a z-score of –2 is two standard deviations below the mean.

References

Bonner, M., Fuchs, M., Maurer, J., Schoenberg, I., and Westheimer, M. (2012). Focus on Grammar (1–5. Pearson Education, White Plains, N.Y., USA).

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

Foley, M. and Hall, D. (2012). MyGrammarLab (1–3. Pearson Education, Harlow, Essex, UK).

Ek, J. and Trim, J (1990) Waystage. Strasbourg, Council of Europe and Cambridge University Press. Available on https://www.coe.int/t/dg4/linguistic/Waystage_CUP.pdf

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

Trim, J. (2009) Breakthrough. Strasbourg, Council of Europe. Available on www.ealta.eu.org

van Ek, J. and Trim, J (1990) Threshold. Strasbourg, Council of Europe, and Cambridge University Press. Available on https://www.coe.int/t/dg4/ linguistic/Threshold-Level_CUP.pdf

van Ek, J. and Trim, J (2001) Vantage. Strasbourg, Council of Europe, and Cambridge University Press. Available on https://www.coe.int/t/dg4/linguistic/Vantage_CUP.pdf


Be yourself
in English.