

# Technical Report: Aligning CEFR–J descriptors to the GSE



October 2016

Mike Mayor  
Daeryong Seo  
John H.A.L. de Jong  
Simon Buckland

# Contents

Executive summary	3
The CEFR and its limitations	3
Introducing the GSE and the GSE Learning Objectives	4
Background to the Common European Framework of Reference – Japan	4
Validating the CEFR-J: background	6
Purpose of this report	7
Overview of the process	8
Results of the study	11
1. Comparing CEFR-J scaling and GSE values for all CEFR-J descriptors	11
2. GSE values of CEFR-J levels	12
3. IRT analysis	13
Conclusions	15
Discussion and future developments	16
References	17
Glossary	18
Appendix 1: Process of development and validation of CEFR-J	19
Appendix 2: Phase II of CEFR-J project	20
Appendix 3: Results of the study	21

## Executive summary

The Global Scale of English (GSE) offers a more granular and detailed means of describing and assessing the progress and performance of English language learners. The GSE Learning Objectives project uses the GSE as the reference scale to extend the 2001 set of Common European Framework of Reference (CEFR) Can Do statements so as to address the needs of more learners, by adding more descriptors for each of the four skills, especially at the lower levels.

In Japan, widespread dissatisfaction with the outcome of English learning in terms of practical communicative ability has led to the adoption of performance-oriented frameworks such as the CEFR for instruction and assessment. The CEFR-J (Common European Framework of Reference – Japan) project adapts the CEFR to Japanese circumstances and requirements, primarily by extending the framework at the lower levels.

The increasing adoption of both CEFR-J and GSE within Japan makes it imperative to assess the alignment between the two systems, and a study has been carried out to assess the degree of correlation. With some caveats, GSE and CEFR-J were found to be highly correlated, and some CEFR-J Can Do statements have been incorporated in slightly modified form into the GSE set.

This study benefits from collaborative support from professors Yukio Tono and Masashi Negishi of Tokyo University of Foreign Studies.

## The CEFR and its limitations

**The CEFR has some imperfections, mainly as regards the range of descriptors available and the consistency of the levels. Both the Global Scale of English and the CEFR-J projects aim to offer remedies.**

As the CEFR was developed with adult learners in Switzerland in mind, it contains no descriptors for learners below A1, and the descriptors are heavily concentrated in A2, B1 and B2, with very few at A1 and the C levels. Moreover, almost two thirds of the descriptors are related to spoken communication, and the levels vary in size: A2, B1 and B2 are twice the size of the other levels, a fact of which learners and teachers are insufficiently aware.

These limitations exclude many learners around the world, including younger Japanese learners, and the Global Scale of English and CEFR-J projects have in common the aim of addressing them.

## Introducing the GSE and the GSE Learning Objectives

**The GSE is a standardised, granular English proficiency scale which runs from 10 to 90, and is psychometrically aligned to the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001).**

**A set of GSE Learning Objectives has been developed, incorporating and extending the CEFR descriptor set. These Learning Objectives have been mapped to the Global Scale of English and describe what a learner can do at different levels of proficiency on the scale.**

Unlike the CEFR and some other scales which describe attainment in broad bands, the Global Scale of English identifies what a learner is likely to be able to do at each point on the scale across speaking, listening, reading and writing skills, to provide a more granular description of increasing language proficiency.

The work to develop the GSE Learning Objectives builds upon and extends the research carried out by Brian North and the Council of Europe in creating the CEFR (North 2000).

The GSE Learning Objectives have been developed by Pearson English over a number of years in collaboration with over 6,000 teachers, ELT authors and language experts from around the world.

## Background to the Common European Framework of Reference – Japan

**The Common European Framework of Reference – Japan (CEFR-J) (TUFS 2012) is an adaptation and extension of the CEFR for English language teaching contexts in Japan. It is composed of Can Do descriptors, following the model of the CEFR, but specifying in more detail the development of English language ability especially at lower levels. It has gone through a series of processes of validation by Japanese experts in English language teaching, partially following the methodology specified in the CEFR.**

The proficiency levels of Japanese EFL learners are among the lowest of all major developed countries, despite various attempts to reform the system of English language teaching in Japan (EF, 2015). The main deficiency of the Japanese system in

place hitherto has been its reliance on instrumental objectives such as passing college entrance examinations. As a result the system has failed (a) to set clear attainment goals for learners as real language users and (b) to motivate them to learn English for communicative purposes and from the perspective of lifelong education.

Since its first publication in 2001, the CEFR has attracted increasing interest in Japan as a potential framework for constructing a solution to this problem, and in 2011 the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) published a report recommending that students' language skills should be assessed using Can Do descriptors – recognising the influence of the CEFR (MEXT 2011). At the same time the Japan Broadcasting Authority (NHK) adopted the CEFR for their foreign language TV and radio programmes.

These and other similar developments led to a project for adapting the CEFR into the Japanese context, specifically for English. A range of surveys of Japanese English learners showed that more than 80% are Non- or Basic Users (in the range from below A1 to A2), with fewer than 20% in the B levels (Independent Users) and almost none in the C levels (Proficient Users) (Tono, 2015).

It was consequently established that further subdivisions would be needed for the lower proficiency levels, following the suggestion in the CEFR document itself for division into sub-levels (Council of Europe 2001, p.23), in order to reflect the specific achievement levels of actual Japanese learners.

The principle aims of the CEFR-J project were accordingly:

1. To subdivide the lower CEFR levels, creating new descriptors to model in more detail the developing proficiency of Japanese learners.
2. To validate empirically a set of new descriptors created for the Japanese context using Japanese experts and following the methodology established in developing the CEFR itself. Given the limited relevance of the C-levels, particularly in Japan, no new descriptors for the C-levels were created, but original CEFR C-level descriptors were adapted for inclusion in the research.

Version 1 of the CEFR-J was published in March 2012, with the following 12 levels:

- Pre-A1
- A1: A1.1, A1.2, A1.3
- A2: A2.1, A2.2
- B1: B1.1, B1.2
- B2: B2.1, B2.2
- C1
- C2

Since its publication the CEFR-J has been widely adopted as a framework for teaching and assessment. It has been downloaded many times by institutes and individuals and has been used as a base framework for creating Can Do descriptors for schools at all levels. Textbooks and dictionaries published in Japan are increasingly including information from the CEFR-J. In the second phase of the project textbook and learner corpora were used to develop grammar, text and learner error profiles.

For an overview of the CEFR-J project to date see Appendices 1 and 2.

## Validating the CEFR-J: background

**The validation exercises carried out during the research and early adoption phase of the CEFR-J were of two types:**

- **Expert rating: teachers and other experts sorted draft descriptors into categories and by level.**
- **Learner self-assessment: learners self-declare how far the descriptors apply to them.**

The results of these two exercises were then combined in the published CEFR-J.

The 'expert rating' exercise was very similar to the procedure followed by Pearson in the development of the GSE Learning Objectives (see De Jong, Mayor and Hayes, 2016). New CEFR-J descriptors were rated alongside existing CEFR descriptors of known value, and teachers' sorting results had extremely positive correlations with the levels defined in the CEFR: 0.93 on average across the four skills (Tono & Negishi, 2012).

The self-assessment exercise was seen as necessary in order to ensure that the descriptors were a close match to learners'

experience of their own communicative skills. Learners were asked to rate each descriptor for ease of performance on a 4-point Likert scale, so that when the findings had been collated the descriptors could be ranked in order of difficulty.

This rank order of descriptors was then subjected to an IRT (Item Response Theory) analysis, as described in the CEFR-J publication (ibid., p. 8) in order to reduce the 'noise' caused by outlying subjects or descriptors. Final adjustments were made, either by changing the wording of some descriptors while maintaining the original order or by re-ordering the descriptors according to the IRT results prior to initial publication in July 2012.

## Purpose of this report

**Both the CEFR-J and GSE projects aim at remedying the imperfections in the CEFR referred to above. They both take the CEFR as a starting point and share the objective of increasing granularity in assessing learner progress, especially at the lower levels. The decision to use the GSE as an additional form of validation for the CEFR-J was therefore logical and consistent with the approach of both projects.**

Through using the GSE it is possible to situate a learner or a learning objective with much greater precision than is possible with the CEFR itself. Moreover, the GSE starts at 10, well below the start of the CEFR at A1 (GSE 22), allowing progress to be described at the very basic level at which, by definition, all learners start.

Similarly, the CEFR-J contains descriptors below A1 for all skills, as well as an enhanced set of descriptors at A1, dividing it into three sub-levels: A1.1, A1.2 and A1.3. A2 is divided into two sub-levels, as suggested in the CEFR itself (Council of Europe 2001, p.23): A2.1 and A2.2. In this way the CEFR-J enables learner progress to be more accurately assessed at these lower levels, where 80% of all Japanese learners are located (Negishi, Takada and Tono 2012).

Since GSE and CEFR-J are measuring essentially the same domain, it is logical and potentially valuable to compare them with each other. The intention of the present project was to create a study which would serve as an additional validation for the CEFR-J, and/or would identify shortcomings or gaps. A further reason for the study was the fact that CEFR-J descriptors had only been rated by Japanese teachers

and learners, whereas GSE Learning Objectives are rated by teachers and experts from all over the world (more than 50 countries). The study would also indicate whether this fact made a difference.

The correlation between the ratings of GSE Learning Objectives and CEFR-J descriptors was found to be quite high ( $r^2 = 0.94$ ). The correspondence was particularly close from A2 to C1 but less at <A1, A1 and C2. This report describes the process of correlating the two scales and its implications for the assessment of learner progress at lower levels.

## Overview of the process

**The CEFR-J descriptors follow the same overall format as the CEFR and GSE, which made it possible to use some of them to generate new GSE Learning Objectives. As part of the ratings process for levelling new GSE Learning Objectives, we also included a number of unchanged CEFR-J descriptors following our standard procedure, in order to scale them to the GSE.**

The comparison of CEFR-J and GSE descriptors was made possible by the fact that they have the same basic structure: 1) performance (task), 2) criteria (quality), 3) condition. For example:

### (CEFR-J)

Descriptor	Performance	Criteria (Quality)	Condition
I can recognise words in a picture book that are already familiar through oral activities.	I can recognise	words in a picture book	that are already familiar through oral activities

### (GSE Learning Objective)

Descriptor	Performance	Criteria (Quality)	Condition
Can recognise a few familiar words, with visual support.	Can recognise	a few familiar words	with visual support

However, there are important differences in the protocol followed by CEFR-J descriptors:

- They start with “I can” rather than “Can”. For example: *I can convey simple information (e.g. times, dates, places), using basic phrases and formulaic expressions.*
- They are often considerably longer than GSE Learning Objectives. For example: *I can actively engage in conversations on a wide range of topics from the general to more specialised*



*cultural and academic fields and express my ideas accurately and fluently.*

- They are frequently ‘portmanteau’, expressing multiple tasks in a single descriptor. For example: *I can give a reasonably smooth presentation about social situations of personal interest, adding my own opinions, and I can take a series of follow up questions from the audience, responding in a way that they can understand.*

This made it necessary to rewrite some of them in order to use them in the study, while others could be used verbatim. The first step was to compare the 108 CEFR-J descriptors with the existing GSE descriptor set in order to identify target descriptors for possible rewriting and rating. As a result of this preliminary examination the CEFR-J descriptors were grouped into three categories:

1. CEFR-J descriptors which differ markedly from GSE in scope and format, and either cover a much wider range of tasks or are educational or metalinguistic in focus. These 49 descriptors were not used in the study. For example: *I can write short texts about my experiences with the use of a dictionary.*
2. CEFR-J descriptors relating to tasks not fully covered in the GSE set. 40 of these in all were adapted and rewritten to fit the GSE format and protocol as potential new GSE descriptors.

For example:

**(CEFR-J original)**

*I can convey very limited information about myself (e.g. name and age), using simple words and basic phrases.*

**(adapted for GSE)**

*Can give very limited personal information using basic fixed phrases.*

In some cases only a portion of a ‘portmanteau’ CEFR-J descriptor was used, in order to ensure focus on a single task. For example:

**(CEFR-J original)**

*I can give a prepared presentation with reasonable fluency, stating reasons for agreement or disagreement or alternative proposals, and I can answer a series of questions.*

**(adapted for GSE)**

*Can agree or disagree with alternative proposals in a discussion, giving reasons.*

3. CEFR-J descriptors which describe identical or similar tasks to GSE Learning Objectives, to be used as 'anchors' in the study, 19 in total. These were rewritten to change 'I can' to 'Can', which it was assumed would not affect their ratings. For example:

**(CEFR-J original descriptor)**

*I can write a series of sentences about my hobbies and likes and dislikes, using simple words and basic expressions.*

**(CEFR-J descriptor rewritten for this study)**

*Can write a series of sentences about hobbies and likes and dislikes, using simple words and basic expressions.*

**(Corresponding GSE descriptor)**

*Can write short texts about their likes and dislikes using basic fixed expressions.*

To summarise the above:

1. 49 CEFR-J descriptors were excluded from the study altogether.
2. 39 CEFR-J descriptors were adapted to be rated as possible new GSE descriptors.
3. 19 CEFR-J descriptors were used verbatim as anchors.

The 39 adapted descriptors were rated alongside 94 GSE Learning Objectives for General Adults, including 22 previously rated anchor items. They were rated by 83 workshop-trained experts and 1416 online participants from over 40 different countries.

The 19 CEFR-J verbatim descriptors were rated alongside a different set of GSE Learning Objectives for General Adults, including the same 22 previously rated anchors and 79 new potential GSE descriptors. 103 trained experts and 923 online raters from over 40 countries took part in this survey, and after removing unreliable and inconsistent raters, 89 expert and 513 online remained: 602 in all. These CEFR-J descriptors were included in an IRT analysis, along with several previous batches of GSE descriptors

# Results of the study

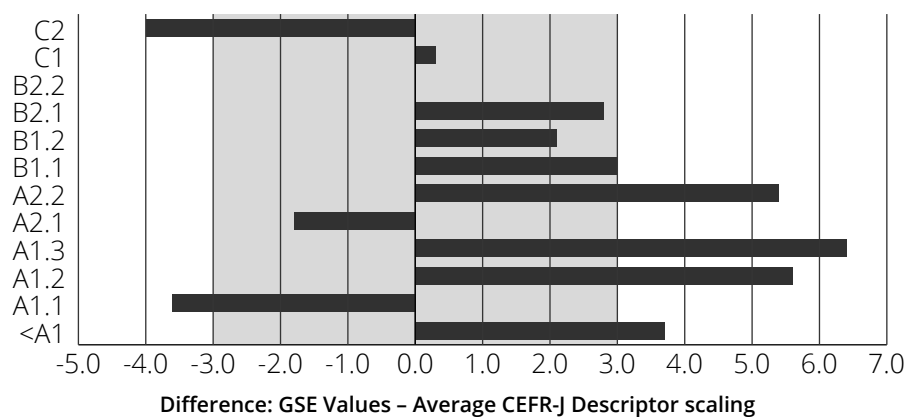
The CEFR-J scaling of the original 58 CEFR-J descriptors was compared with their GSE value resulting from the current study. The average GSE values for the CEFR-J levels were computed. For the total set of 58 CEFR-J descriptors, 86% of the variance in GSE values could be explained ( $r^2$ ) by their original CEFR-J scaling in a linear regression. This is in fact an underestimate of the true relation between the CEFR-J and the GSE since 97% of the variance in the 19 CEFR-J descriptors used verbatim was explained by the GSE.

Three analyses were carried out, and summarised below. For full results see Appendix 3.

## 1. Comparing CEFR-J scaling and GSE values for all CEFR-J descriptors

All CEFR-J descriptors (anchors and adapted) were rated on GSE according to the standard procedures of Pearson’s GSE research (by experts and online respondents, see De Jong, Mayor and Hayes, 2016) and compared to their original CEFR-J levels. Descriptors for which the CEFR-J scaling fell outside of the GSE interval for the corresponding CEFR level were flagged (see Appendix 3).

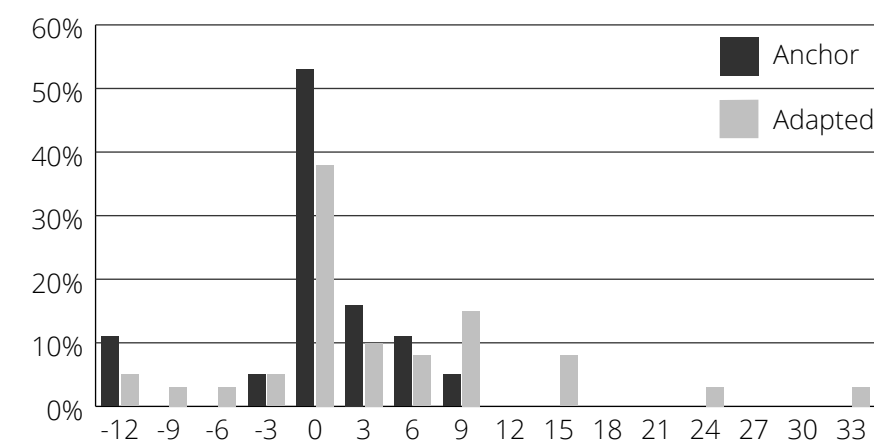
Figure 1 shows the average difference between GSE values found in this study and the closest CEFR-J cut-off by level. If the difference for a level is 0 (zero) it means that the GSE value falls within the interval on the GSE for that CEFR-J level. The shaded area shows a tolerance of 3 GSE points around a zero difference. The average descriptor in two out of the 12 CEFR-J levels (C2 and A1.1) is significantly more difficult than was found in the current study. It must be noted, however, that the discrepancy at C2 might have been caused by the adaptations made to the C-level descriptors for the purpose of this study. By contrast, descriptors from four CEFR-J levels (A2.2, A1.3, A1.2 and <A1) fall on average significantly below the GSE value.



**FIGURE 1**  
Average difference between GSE values found in this study and the closest CEFR-J cut-off by level

Looking at individual descriptors, we must distinguish between the unmodified (verbatim) anchor descriptors and the descriptors that were adapted for this study, because any differences in level for these descriptors could have been caused by their modification. Figure 2 shows the difference per descriptor between the GSE value obtained in this study and the closest CEFR-J cut-off.

More than 50% of the anchor descriptors fall within the -3 to +3 tolerance interval. As might be expected, Figure 2 shows that the size of the differences is larger for the adapted descriptors than for the anchor descriptors.



**FIGURE 2**  
Difference per descriptor between the GSE value obtained in this study and the closest CEFR-J cut-off

2. GSE values of CEFR-J levels

GSE values were computed for all CEFR-J levels. The following table compares the results with the standard GSE equivalents for the CEFR levels. The interval on the GSE corresponding to CEFR A1 has been divided in three equal parts to mirror the CEFR-J subdivision of A1.

CEFR-J	Pre-A1	A1.1	A1.2	A1.3	A2.1	A2.2	B1.1	B1.2	B2.1	B2.2	C1	C2
GSE value	20	20	31	36	30	43	52	58	65	72	81	81
CEFR	Tourist	A1			A2	A2+	B1	B1+	B2	B2+	C1	C2
GSE range	13–21	22–24	25–26	27–29	30–36	37–42	43–50	51–58	59–67	68–75	76–84	85–90

This suggests that most of the CEFR-J levels are generally ordered in agreement with the CEFR, but that there are some discrepancies. These may of course be due to the adaptation of two thirds of the descriptors in the comparison, but limiting this comparison to the 19 anchor descriptors would constitute a non-representative sample for computing reliable averages for twelve levels.

**TABLE 1**  
CEFR-J levels and sub-levels compared with their GSE equivalents

### 3. IRT analysis

It was judged that the extent of rewording involved in the adapted CEFR-J descriptors made any comparison between the original CEFR-J level and GSE values no more than indicative. Only the 19 unchanged CEFR-J descriptors anchors were therefore included in the IRT analysis which was subsequently carried out on all GSE descriptors rated in three previous rating exercises.

The purpose of the IRT analysis was to bring together all the data from 20 sets of learning objectives (for the four different audiences of learners: Adult Learners, Young Learners, Learners of Academic English, Learners of Professional English, as well as CEFR-J) and ensure that they were all calibrated to the same single scale so that scores could be directly compared.

The cleaned data sets were combined to create a file of 463 GSE Learning Objectives and 1,449 raters. The software programme WINSTEPS (Linacre, 1998; 2005) was used to perform the analysis; after experimentation in a pilot stage, a free-calibration 1-parameter (i.e., Rasch) model was decided on as psychometrically optimal and technically feasible.

The data was calibrated and scaled by a four-stage equating: 1) free calibration, 2) stability check, 3) recalibration of difficulty estimates of new items using already-existing item parameters (i.e., fixed method, Taherbhai & Seo, 2007; Taherbhai & Seo, 2013), and 4) application of regression coefficients and GSE scaling constants. At the first stage, the following criteria were applied to eliminate raters and learning objectives:

- Too few observations to be representative of the world community of teachers (i.e.  $N < 80$ )
- The ratings did not fit the chosen 1-parameter model (i.e. the INMSQ and/or the OUTMSQ value for the rater or the GSE Learning Objective was  $> 2.56$ )
- The rater rated fewer than 25 Learning Objectives or their rating had a point biserial of  $< 0.10$
- The frequency distribution of the ratings for a GSE Learning Objective showed an irregular pattern (i.e. several outlier responses)

After the first stage, 11 learning objectives were eliminated from the data file. At the second stage, a stability check was performed on an anchor pool (i.e., a pool of 32 anchor items) using a method proposed by Taherbhai and Seo (2007; 2013). The stability check identified 2 learning objectives, and all descriptors were recalibrated by fixing each surviving anchor learning objective with its item difficulty obtained from the

previous IRT analysis to place the current objectives on the same scale (at Stage 3).

At the final stage, we applied regression constants on all the current descriptors to place them onto the North 2000 scale and then applied the scaling formula into the transformed item difficulty of each objective to obtain the final GSE values.

The limited size of the descriptor set in this study does not allow us to estimate the cut-offs between CEFR-J levels on the GSE. Assuming however that descriptors are representative of the CEFR-J levels to which they are assigned, we can compare these averages to the midpoints on the GSE intervals for the corresponding CEFR levels. Figure 3 shows the midpoints of the CEFR intervals on the GSE plotted against the average GSE value of CEFR-J descriptors per level.

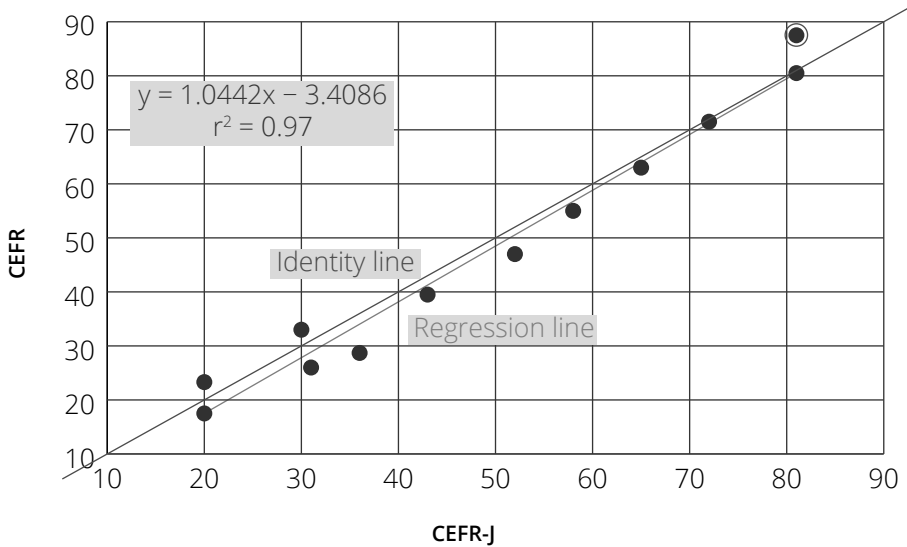


Figure 3 clearly provides evidence of the high level of correspondence between the two measurement scales with 97% of the variance explained in a simple linear regression model. The only outlier is level C2 (marked with a red circle) which is significantly underestimated in the CEFR-J. Given the distribution of English proficiency levels among Japanese students with the vast majority being at the lowest levels, this lack of precision at the C2 level may not constitute a serious drawback.

**FIGURE 3**  
Correspondence between CEFR-J and GSE

## **Conclusions**

- GSE ratings of most individual CEFR-J descriptors corresponded well to their CEFR-J scaling.
- The variance of the averages of CEFR-J descriptors per level explains 97% of the variance of CEFR levels on the GSE.
- More data on the A and Pre-A level descriptors would be needed to satisfactorily support the CEFR-J division of levels within this range.

Finally, a word of caution: conclusions to be drawn from the IRT analysis were limited by the lack of any direct comparison between the original Japanese data and the new (GSE) data sets. The comparison was made based on the already analysed and interpreted Japanese data, not the original (raw) data which the CEFR-J team had worked from.

## Discussion and future developments

There are a number of explanations for the discrepancies noted between some CEFR-J levels and GSE scores, especially regarding the CEFR-J sublevels:

1. A relatively small number of descriptors per level were available in the data set. With such small data sets outliers cannot be identified and excluded from the results.
2. There was also a difference in methodology: learner self-assessment was one of the data sets in the Japanese analysis, while only expert (teacher) ratings were used for the GSE. It is highly likely that these two different methodologies could explain some of the differences in the results.
3. The use of self-assessment may also explain why CEFR-J ratings are systematically at the top of or higher than the GSE ranges. Asian learners self-assessing tend to underrate their abilities.

We look forward to the opportunity for further collaboration with the CEFR-J team and to running more studies to evaluate additional descriptors with the aim of refining the alignment of the A1 sublevels, and correcting or adjusting some of the other divergences. There is also scope to collaborate on the alignment of our grammar and vocabulary projects.

In conclusion, a clear and systematic alignment between GSE and CEFR-J has been demonstrated, both as regards the rank order of descriptors and their scores on GSE. The divergences noted, especially on the A1 sublevels, may be explained by the relatively small amount of data available.



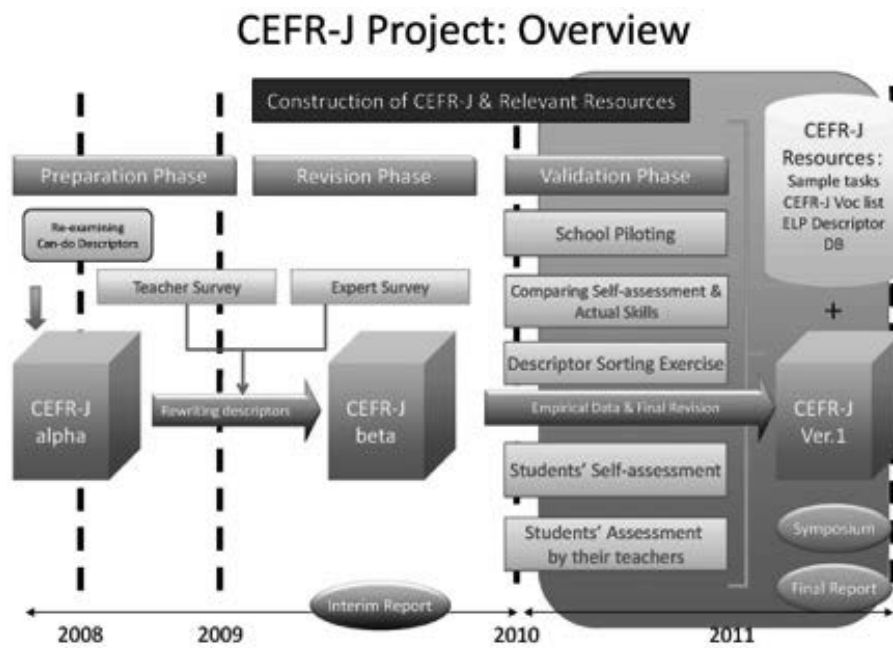
## References

- 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://www.english.com/gse/researchers>
- EF EPI: EF English Proficiency Index (5th edition). Education First, 2015. Available at [http://media2.ef.com/\\_/~/media/centraleftcom/epi/downloads/full-reports/v5/ef-epi-2015-english.pdf](http://media2.ef.com/_/~/media/centraleftcom/epi/downloads/full-reports/v5/ef-epi-2015-english.pdf)
- Ministry of Education, Culture, Sports, Science and Technology – Japan (MEXT) (2011) *Five Proposals and Specific Measures for Developing Proficiency in English for International Communication*. Accessible at <http://www.mext.go.jp/english/elsec/1319701.htm>.
- Negishi, M., Takada, T. and Tono, Y. (2012) A progress report on the development of the CEFR-J. *Studies in Language Testing* 36: 137-165. Cambridge: Cambridge University Press.
- North, B. (2000) *The development of a common framework scale of language proficiency*. New York: Peter Lang.
- Taherbhai, H., & Seo, D. (2007). Comparing concurrent versus fixed parameter equating with common items, using the one-parameter Rasch and the Rasch partial credit model in a mixed-item format test, *Journal of Applied Measurement*, 8, 84-96
- Taherbhai, H., & Seo, D. (2013). The philosophical aspects of IRT equating: modeling drift to evaluate cohort growth in large scale assessments. *Educational Measurement: Issues and Practice*, 32, 2-14.
- Tono, Y., Negishi, M. (2012) *Adapting the CEFR for English Language Teaching in Japan*. Tokyo University of Foreign Studies. JALT FLP SIG Newsletter No.8: 5-12. Accessible at [http://www.tufts.ac.jp/ts/personal/tonolab/cefr-j/Tono&Negishi2012forJALT\\_FLP SIG.pdf](http://www.tufts.ac.jp/ts/personal/tonolab/cefr-j/Tono&Negishi2012forJALT_FLP SIG.pdf)
- Tono, Y. (2015) From CAN-DO to Classroom: How to pedagogically contextualize the CEFR-J. JALT FLP SIG: Second Conference on Critical constructive assessment of CEFR-based language teaching in Japan and beyond.
- TUFS (2016). *A Symposium on the CEFR-J RLD Project*. AGORA GLOBAL, Tokyo University of Foreign Studies (TUFS).

## Glossary

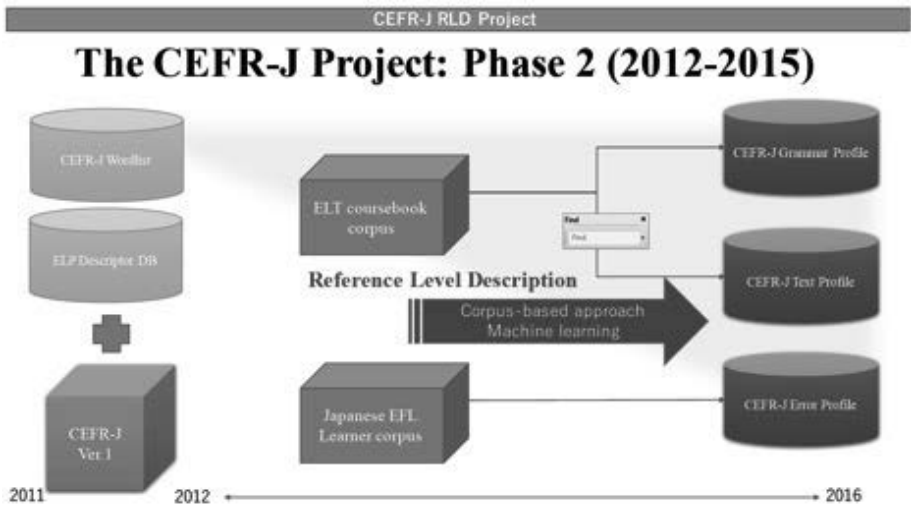
Term	Definition
anchor item	A test item or learning objective which has a known difficulty value from earlier research. It is used to link new items to the same scale.
correlation	A statistic showing the interdependence between two variables
explained variance ( $r^2$ )	A statistical measure of how close the data are to the fitted regression line - in other words, how accurately the variances in the data have been modelled. An $r^2$ of 1 denotes a perfect fit; any value above 0.7 is usually regarded as satisfactory.
GSE Learning Objective	A description of what a student can do at a particular point on the Global Scale of English. Learning objectives are also referred to as "Can Do statements" and "descriptors"
INMSQ and OUTMSQ	Infit and outfit mean square: two statistics used in IRT to show how well the data fits the model
IRT	Item Response Theory: A method used to apply a mathematical model to test data. It predicts the probability of a test item being correctly answered based on the ability of the person and the difficulty of the item.
point biserial	A statistic showing the correlation between an item (dichotomous i.e., correct/incorrect) contributing to a scale and a continuous score scale.
portmanteau descriptor	One which covers more than one task or skill (e.g. asking and answering questions, or reading and summarising a text)

## Appendix 1: Process of development and validation of CEFR-J



Source: TUFFS (2016)

## Appendix 2: Phase II of CEFR-J project



Source: TUFS (2016)

## Appendix 3: Results of the study

Adapted descriptor as rated	Status	Skill	GSE	CEFR	CEFR-J level	CEFR-J original descriptor
Can recognise a few familiar everyday words, if delivered slowly and clearly.	ADAPTED	Listening	10	<A1	PreA1	I can catch everyday, familiar words, provided they are delivered clearly and slowly.
Can recognise the letters of the English alphabet when pronounced.	ADAPTED	Listening	11	<A1	PreA1	I can catch key letters of the English alphabet, when they are pronounced.
Can recognise a few familiar words, with visual support.	ADAPTED	Reading	13	<A1	PreA1	I can recognise words in a picture book that are already familiar through oral activities.
Can give very limited personal information using basic fixed phrases.	ADAPTED	Spoken Production   Speaking	16	<A1	PreA1	I can convey very limited information about myself (e.g. name and age), using simple words and basic phrases.
Can follow instructions for everyday tasks, if delivered slowly and clearly, with visual support.	ADAPTED	Listening	24	A1	A2.2	I can understand the instructions about procedures (e.g. cooking, handicrafts), with visual aids, provided they are delivered in slow and clear speech involving rephrasing and repetition.
Can express basic likes and dislikes using basic fixed expressions.	ADAPTED	Spoken Production   Speaking	26	A1	A1.1	I can exchange simple opinions about very familiar topics such as likes and dislikes for sports, foods, etc., using a limited repertoire of expressions, provided that people speak clearly.
Can introduce themselves, their hobbies and interests in a basic way.	ADAPTED	Spoken Production   Speaking	26	A1	A2.1	I can introduce myself including my hobbies and abilities, using a series of simple phrases and sentences.
Can follow instructions for everyday tasks, if delivered slowly and clearly.	ADAPTED	Listening	26	A1	A2.2	I can understand and follow a series of instructions for sports, cooking, etc. provided they are delivered slowly and clearly.

Can accept offers using basic fixed expressions.	ADAPTED	Spoken Interaction   Speaking	27	A1	A1.2	I can make, accept and decline offers, using simple words and a limited range of expressions.
Can make requests related to immediate needs using basic fixed expressions.	ADAPTED	Spoken Production   Speaking	30	A2	PreA1	I can express my wishes and make requests in areas of immediate need such as "Help!" and "I want", using basic phrases. I can express what I want by pointing at it, if necessary.
Can give simple descriptions of everyday objects using basic fixed expressions.	ADAPTED	Speaking	31	A2	A1.1	I can give simple descriptions e.g. of everyday object, using simple words and basic phrases in a restricted range of sentence structures, provided I can prepare my speech in advance.
Can make basic offers using fixed expressions.	ADAPTED	Speaking	32	A2	A1.2	I can make, accept and decline offers, using simple words and a limited range of expressions.
Can get the gist of short simple narratives, with visual support.	ADAPTED	Reading	32	A2	A1.3	I can understand short narratives with illustrations and pictures written in simple words.
Can decline offers using basic fixed expressions.	ADAPTED	Spoken Interaction   Speaking	33	A2	A1.2	I can make, accept and decline offers, using simple words and a limited range of expressions.
Can write short texts about their likes and dislikes using basic fixed expressions.	ADAPTED	Writing	34	A2	A1.2	I can write short texts about matters of personal relevance (e.g. likes and dislikes, family, and school life), using simple words and basic expressions.
Can give simple opinions using basic fixed expressions..	ADAPTED	Speaking	34	A2	A1.3	I can get across basic information and exchange simple opinions, using pictures or objects to help me.

Can express wishes in areas of immediate need using basic fixed expressions.	ADAPTED	Spoken Production   Speaking	36	A2+	PreA1	I can express my wishes and make requests in areas of immediate need such as "Help!" and "I want!", using basic phrases. I can express what I want by pointing at it, if necessary.
Can understand short, simple narratives and biographies.	ADAPTED	Reading	39	A2+	n/a	modified from previous descriptor
Can write descriptions of everyday personal experiences.	ADAPTED	Writing	41	A2+	A2.2	I can write a simple description about events of my immediate environment, hobby, places, and work, provided they are in the field of my personal experience and of my immediate need.
Can give a short talk about a familiar topic, with visual support (e.g. photos).	ADAPTED	Speaking	44	B1	A1.3	I can give a brief talk about familiar topics (e.g. my school and my neighborhood) supported by visual aids such as photos, pictures, and maps, using a series of simple phrases and sentences.
Can get the gist of explanations of unfamiliar cultural practices and customs if delivered slowly and clearly.	ADAPTED	Listening	48	B1	B1.1	I can understand the gist of explanations of cultural practices and customs that are unfamiliar to me, provided they are delivered in slow and clear speech involving rephrasing and repetition.
Can give opinions about news and stories using basic fixed expressions.	ADAPTED	Writing	49	B1	A2.2	I can write my impressions and opinions briefly about what I have listened to and read (e.g. explanations about lifestyles and culture, stories), using basic everyday vocabulary and expressions.

Can express hopes for the future using a range of fixed expressions.	ADAPTED	Speaking	51	B1+	B1.1	I can express opinions and exchange information about familiar topics (e.g. school, hobbies, hopes for the future), using a wide range of simple English.
Can discuss the main points of news stories about familiar topics.	ADAPTED	Spoken Interaction   Speaking	52	B1+	B2.1	I can discuss the main points of news stories I have read about in the newspapers/ on the internet or watched on TV, provided the topic is reasonably familiar to me.
Can write personal emails or letters, reporting recent events in detail.	ADAPTED	Writing	55	B1+	B1.2	I can write narratives (e.g. travel diaries, personal histories, personal anecdotes) in several paragraphs, following the order of events. I can write personal letters which report recent events in some detail.
Can describe an everyday consumer-related problem and request a correction or solution.	ADAPTED	Speaking	60	B2	B1.1	I can explain with confidence a problem which has arisen in familiar places such as a station or a shop (e.g. purchasing the wrong ticket) and obtain the right product or service by requesting politely and expressing gratitude (assuming that the provider of the service is cooperative).
Can write an extended description of events taking place in their immediate environment.	ADAPTED	Writing	61	B2	B1.1	I can write a description of substantial length about events taking place in my immediate environment (e.g. school, workplace, local area), using familiar vocabulary and grammar.
Can understand the plot of extended narratives written in standard, non-literary language.	ADAPTED	Reading	62	B2	B1.2	I can understand the plot of extended narratives written in plain English.



Can recommend a course of action, giving reasons.	ADAPTED	Spoken Interaction   Speaking	62	B2	B2.1	I can give a prepared presentation with reasonable fluency, stating reasons for agreement or disagreement or alternative proposals, and I can answer a series of questions.
Can express opinions about news stories using a wide range of everyday language.	ADAPTED	Spoken Interaction   Speaking	65	B2	B2.1	I can discuss the main points of news stories I have read about in the newspapers/ on the internet or watched on TV, provided the topic is reasonably familiar to me.
Can agree or disagree with alternative proposals in a discussion, giving reasons.	ADAPTED	Spoken Interaction   Speaking	65	B2	B2.1	I can give a prepared presentation with reasonable fluency, stating reasons for agreement or disagreement or alternative proposals, and I can answer a series of questions.
Can give an extended talk or presentation on a non-technical topic, with visual support.	ADAPTED	Spoken Production   Speaking	67	B2+	A2.2	I can make a short speech on topics directly related to my everyday life (e.g. myself, my school, my neighborhood) with the use of visual aids such as photos, pictures, and maps, using a series of simple words and phrases and sentences.
Can understand differences and similarities between different points of view in extended texts.	ADAPTED	Reading	67	B2+	B2.1	I can read texts dealing with topics of general interest, such as current affairs, without consulting a dictionary, and I can compare differences and similarities between multiple points of view.
Can put forward a smoothly flowing and logical structured argument, highlighting significant points.	ADAPTED	Spoken Interaction   Speaking	73	B2+	C2	I can present a clear, smoothly flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.

Can get the gist of specialised articles and technical texts outside their field.	ADAPTED	Reading	74	B2+	C1	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.
Can write essays and reports synthesising information from a number of sources.	ADAPTED	Writing	76	C1	B2.1	I can write reasonably coherent essays and reports using a wide range of vocabulary and complex sentence structures, synthesising information and arguments from a number of sources, provided I know something about the topics.
Can write summaries and reviews of professional or literary works.	ADAPTED	Writing	80	C1	C2	I can write clear, smoothly flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.
Can write about complex subjects, underlining the salient issues and in a style suited to the intended reader.	ADAPTED	Writing	85	C2	C1	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.

Can write linguistically complex and logically structured reports and articles.	ADAPTED	Writing	85	C2	C2	I can write clear, smoothly flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.
Can understand long and linguistically complex factual and literary texts, appreciating distinctions of style.	ADAPTED	Reading	87	C2	C1	I can understand long and complex factual and literary texts, appreciating distinctions of style.
Can understand short, simple instructions such as "Stand up." "Sit down." "Stop." etc., provided they are delivered face-to-face, slowly and clearly.	ANCHOR	Listening	10	<A1	A1.1	I can understand short, simple instructions such as "Stand up." "Sit down." "Stop." etc., provided they are delivered face-to-face, slowly and clearly.
Can understand a fast-food restaurant menu that has pictures or photos, and choose the food and drink in the menu.	ANCHOR	Reading	13	<A1	A1.1	I can understand a fast-food restaurant menu that has pictures or photos, and choose the food and drink in the menu.
Can express wishes and make requests in areas of immediate need such as "Help!" and "I want", using basic phrases.	ANCHOR	speaking	24	A1	PreA1	I can express wishes and make requests in areas of immediate need such as "Help!" and "I want", using basic phrases.
Can respond simply in basic, everyday interactions such as talking about what can/cannot do or describing colour, using a limited repertoire of expressions.	ANCHOR	speaking	29	A1	A1.2	I can respond simply in basic, everyday interactions such as talking about what I can/I cannot do or describing colour, using a limited repertoire of expressions.
Can write a series of sentences about hobbies and likes and dislikes, using simple words and basic expressions.	ANCHOR	Writing	33	A2	A1.2	I can write a series of sentences about my hobbies and likes and dislikes, using simple words and basic expressions.

Can understand phrases and expressions related to matters of immediate relevance to self or family, school, neighbourhood etc., provided they are delivered slowly and clearly.	ANCHOR	Listening	33	A2	A1.3	I can understand phrases and expressions related to matters of immediate relevance to myself or my family, school, neighborhood etc., provided they are delivered slowly and clearly.
Can give simple directions from place to place, using basic expressions such as “turn right” and “go straight” along with sequencers such as first, then, and next.	ANCHOR	speaking	33	A2	A2.1	I can give simple directions from place to place, using basic expressions such as “turn right” and “go straight” along with sequencers such as first, then, and next.
Can understand and follow a series of instructions for sports, cooking, etc. provided they are delivered slowly and clearly.	ANCHOR	Listening	40	A2+	A2.2	I can understand and follow a series of instructions for sports, cooking, etc. provided they are delivered slowly and clearly.
Can understand the clearly written instructions (e.g. for playing games, for filling in a form, for assembling things).	ANCHOR	Reading	44	B1	B1.1	I can understand clearly written instructions (e.g. for playing games, for filling in a form, for assembling things).
Can understand the gist of explanations of cultural practices and customs that are unfamiliar to me, provided they are delivered in slow and clear speech involving rephrasing and repetition.	ANCHOR	Listening	48	B1	B1.1	I can understand the gist of explanations of cultural practices and customs that are unfamiliar to me, provided they are delivered in slow and clear speech involving rephrasing and repetition.
Can write impressions and opinions briefly about what has been listened to or read (e.g. explanations about lifestyles and culture, stories), using basic everyday vocabulary and expressions.	ANCHOR	Writing	51	B1+	A2.2	I can write impressions and opinions briefly about what I have listened to or read (e.g. explanations about lifestyles and culture, stories), using basic everyday vocabulary and expressions.
Can write personal letters which report recent events in some detail.	ANCHOR	Writing	51	B1+	B1.2	I can write personal letters which report recent events in some detail.

Can write narratives (e.g. travel diaries, personal histories, personal anecdotes) in several paragraphs, following the order of events.	ANCHOR	Writing	61	B2	B1.2	I can write narratives (e.g. travel diaries, personal histories, personal anecdotes) in several paragraphs, following the order of events.
Can get the right treatment by providing relevant, detailed information.	ANCHOR	speaking	63	B2	B1.2	I can get the right treatment by providing relevant, detailed information.
Can recognise necessary information and the points of the argument from articles and reference materials in their specialised field without consulting a dictionary.	ANCHOR	Reading	72	B2+	B2.2	I can recognise necessary information and the points of the argument from articles and reference materials in my specialised field without consulting a dictionary.
Can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly.	ANCHOR	Listening	78	C1	C1	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly.
Can understand specialised articles and longer technical instructions, even when they do not relate to their field.	ANCHOR	Reading	80	C1	C1	I can understand specialised articles and longer technical instructions, even when they do not relate to my field.
Can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points.	ANCHOR	Writing	80	C1	C2	I can write complex letters, reports or articles which present a case with an effective logical structure which helps the other person to notice and remember significant points.
Can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms.	ANCHOR	speaking	85	C2	C2	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms.

Be yourself  
in English.