Pearson

Research Series

Technical Report

Aligning EIKEN Descriptors to the GSE

September 2017

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

Global Scale of English
Fast-track your progress

Contents

Executive summary

The EIKEN Test in Practical English Proficiency is one of the most widely used English-language testing programs in Japan, recognized for educational and professional uses within and outside Japan. The need for international recognition makes it desirable to align EIKEN levels to other widely recognised measurement scales, such as the Common European Framework (CEFR) and the Global Scale of English (GSE). The greater granularity of the GSE and the much larger number of descriptors (Learning Objectives) makes it particularly suitable for this purpose.

EIKEN descriptors follow a similar format to those of the CEFR and the GSE, so they were suitable for rating alongside GSE Learning Objectives in a standard rating exercise. Between 100 and 190 raters' responses were collected for a set of 63 EIKEN descriptors along with 385 Pearson descriptors. The EIKEN set was analysed along with 385 new GSE Learning Objectives according to the Rasch rating scale model, as used when developing all GSE Learning Objectives.

High agreement was found between EIKEN levels and the GSE, with 78% of GSE variance explained by EIKEN descriptors.

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 (LOs) 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 for speaking, listening, reading and writing skills, providing 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 North and the Council of Europe in creating the CEFR (North 2000). The GSE Learning Objectives have been developed by Pearson over several years in collaboration with over 6,000 teachers, ELT authors and language experts from around the world.

Background to the EIKEN levels and their specifications

The EIKEN Test in Practical English Proficiency is one of the most widely used English language testing programs in Japan. It is offered at 7 levels, and more than 95 million learners have taken the test since its inception in 1963. EIKEN is recognized for a very broad range of educational and professional uses within Japan and also as a language skills certificate for study abroad. It is also accepted by many universities and educational institutions in North America, Australia, and throughout the world.

There are seven tests within the EIKEN framework, each representing a different ability level. The levels are called grades and are awarded on a pass-or-fail basis.

Each EIKEN grade is assessed by a separate test given in two stages, both of which must be passed: (1) a paper-based test that evaluates reading, listening, and writing, and (2) a speaking test in the form of a face-to-face interview or online spoken assessment. Test-takers are given separate scores for reading, listening, speaking, and writing, plus a composite score.

The EIKEN tests report on a specially-developed scale called the CSE (Common Scale for English), which has been aligned to the CEFR (Dunlea & Matsudaira, 2009) and to other international standards such as TOEFL, TOEIC, etc.
Each of the seven EIKEN levels is additionally defined by a series of Can-Do statements which exist in Japanese and in English, such as:

- *Can understand practical texts describing how to do things.* (Reading, Grade 2)
- *Can write about his/her hobbies or interests.* (Writing, Grade 3)

Purpose of validating EIKEN descriptors against the GSE

Although the process of seeking international recognition for the EIKEN levels goes back to 2002, the CSE (Common Scale of English) now used to report EIKEN levels is of recent adoption. The validation and alignment studies carried out so far on the CSE and EIKEN levels have been mostly specification-based. The sets of Can-Do statements corresponding to the different EIKEN levels are derived from self-assessment questionnaires from 20,000 Japanese test takers. There is therefore a strong case for further comparative studies of EIKEN levels against other measurement scales for proficiency in English.

The only published study to date on EIKEN-CEFR alignment (Dunlea & Matsudaira, 2009) was based on content analysis and comparison of real-world benchmarks such as college admission, and covered Grades 1 and Pre-1 (C1 and B2) only. There was therefore a need for a study comparing all EIKEN levels with an independent set of learning objectives such as the present study using the GSE.

The following considerations are relevant to the current study:

- The GSE itself is linearly aligned with the CEFR; such a study could therefore support the existing alignment study of CEFR and EIKEN levels.
- Pearson have added over 450 new descriptors to the original CEFR descriptor set, as well as new sets of descriptors for Professional, Academic and Young Learners. All of these have been validated using a similar methodology to that employed in North's original study (North, 2000; De Jong, Mayor, & Hayes, 2017). Availability of so many new descriptors offers substantially more reference points for aligning any test or test framework to the CEFR.
- Both the CEFR and EIKEN are level-based frameworks. It is therefore unlikely that the CEFR and EIKEN levels would line up exactly. The higher level of granularity of the GSE offers greater opportunity to establish the positions of the Eiken levels cutoffs.

The goal of the present study was to validate the alignment of EIKEN and CEFR levels. It should be noted that the positioning of the EIKEN and GSE Learning Objectives was obtained using different methodologies: self-assessment by learners in the case of EIKEN vs. standard-setting by experts in the case of the GSE (assessing new descriptors alongside anchor descriptors of known value).

EIKEN, GSE and CEFR descriptors

The EIKEN descriptors follow a similar format to the CEFR and GSE descriptors, which made it possible to rate some of them alongside new GSE Learning Objectives in a rating exercise. A number of previously rated descriptors with known values from both the CEFR and the GSE were included as anchors.

Many EIKEN and GSE descriptors have the same basic structure: they consist of (a) a performance or action, (b) a target or object, and (c) a specification of quality and/or the conditions under which the task is performed. In addition, the EIKEN descriptors usually provide examples of the targets or objects in question.

The following similar descriptors from EIKEN and from GSE illustrate this.

	Descriptor	Task	Target	Condition/ quality	LEVEL
Eiken	can say what he/she likes and dislikes and explain in simple terms why (e.g. animals, food, sports)	can say … and explain	what he/she likes and dislikes, and … why (e.g. animals, food, sports)	in simple terms	grade 3 (A1)
GSE	can describe a person's likes and dislikes using simple language. (P)	can describe	a person's likes and dislikes	using simple language	28 (A1)

However, a large number of EIKEN descriptors, especially those at a higher level, specify only the task and target, while the comparable GSE descriptor also includes a quality specification. For example:

	Descriptor	Task	Target	Condition/ quality	LEVEL
Eiken	can understand instruction manuals for products (e.g. electrical appliances)	can understand	instruction manuals for products (e.g. electrical appliances).		grade Pre-1 (B2)
GSE	can understand clearly written, straightforward instructions on how to use a piece of equipment. (CA)	can understand	instructions on how to use a piece of equipment	clearly written, straightforward	46 (B1)

There were, however, enough cases where either parallel descriptors from EIKEN and GSE both had a condition or quality specification, or where neither did, to enable the selection of 63 EIKEN descriptors for rating alongside new GSE descriptors.

For example (parallel EIKEN and GSE descriptors mentioning a condition or a quality specification):

	Descriptor	Task	Target	Condition/ quality	LEVEL
Eiken	can understand a speaker on the telephone, provided the content is simple (e.g. agreeing when to meet, taking short messages)	can understand	a speaker	on the telephone, provided the content is simple	grade Pre-2 (A2)
GSE	can understand simple work-related questions asked on phone calls. (P)	can understand	work-related questions	simple, asked on phone calls	39 (A2)

Or, an example not mentioning a condition or quality specification:

	Descriptor	Task	Target	Condition/ quality	LEVEL
Eiken	can talk about dreams and ambitions for his/her future (e.g. countries that he/she would like to visit, the career he/she would like to pursue)	can talk about	dreams and ambitions for his/her future		grade Pre-2 (A2)
GSE	can describe dreams, hopes and ambitions. (C)	can describe	dreams, hopes and ambitions		48 (B1)

For every one of the selected EIKEN descriptors, the examples given were equally relevant also to the GSE/CEFR descriptor, making them sufficiently comparable for the purposes of this study.

The rating process

The combined set of EIKEN and GSE learning objectives went through the same rating process as earlier batches of GSE Learning Objectives.

Two groups of raters were involved:

1. 'Expert' raters: knowledgeable about the CEFR, and with experience in teaching and/or curriculum design. This group was trained on the Global Scale of English and given a standardisation exercise.
2. 'Online' raters: teachers with substantial experience and at least some familiarity with the CEFR.

The standard rating methodology presents both groups of raters with a set of newly developed GSE Learning Objectives, together with around 20% of anchor items, mostly taken from North's original research (2000), on which the CEFR itself was based. In this research, North placed descriptors on the highly granular logit scale that results from applying Item Response Theory (IRT). North applied the one-parameter Rasch Model (Rasch, 1960), which yields a one-dimensional scale ranging theoretically from minus infinity to plus infinity, but in practice, depending on the data, is usually limited to values between -6 and 6. North obtained a scale ranging from –5.68 to 4.68. For reporting purposes he divided this scale in regular intervals of approximately one logit. On publication of the CEFR, these intervals were applied to create the CEFR levels.

The Global Scale of English is a linear transformation of the original North scale, making it possible for users to order CEFR descriptors by difficulty within a CEFR level. The North anchor items therefore had known values on the GSE (and within a CEFR level), which made it possible to rank the EIKEN and other new descriptors with reference to them once they had all been rated. For more detailed information on developing and scaling descriptors for the CEFR, please refer to Appendix A of the original CEFR publication: Council of Europe, 2001, 205–216; for more information on the GSE and its derivation from the original North scale, refer to De Jong, Hayes and Mayor, 2016.

Each set of descriptors numbered about 100, covered all four skills as well as a range of predicted CEFR levels. Descriptors were presented to raters by skill, in a random order. Expert raters rated all descriptors in the set and directly assigned GSE values, while for the Online group the set was subdivided into overlapping subsets (including anchor items, also overlapping) of 40 descriptors. Raters assigned a CEFR level to each descriptor.

The data collected went through a three step analysis process:

1. In the first stage, data for each set of descriptors went through classical analysis to identify and remove highly incongruent data points. (Please refer to the Certainty index in Appendix 2). Degree of agreement among raters rating a particular descriptor, defined as the proportion of raters assigning values within the same interval on the scale is expressed as a certainty value with a minimum of zero and maximum of 1 was calculated for each descriptor. Certainty values below 0.7 were flagged.
2. The second stage was an IRT analysis based on the Rasch model (Rasch, 1960) using the computer program WINSTEPS (Linacre, 2009) which placed all GSE descriptors from different data sets on a single scale.
3. The third stage was a final review of the data and content using a checklist.

The 63 selected EIKEN descriptors were analysed alongside 319 new GSE descriptors and 66 anchors (some GSE and some CEFR originals) from previous batches. These anchors were not the same ones as those matching the EIKEN descriptors. Each EIKEN descriptor was rated by between 100 and 190 raters and was analysed along with the GSE descriptor. It should be noted that raters were not given any information as to the origin or provenance of the descriptors.

Results of the Study

The aim of the study was to validate statistically the extent to which EIKEN levels and learning objectives are aligned to the GSE. Three analyses were carried out using a Rasch rating scale model, and the results showed a good correlation, with certainty >0.7 in all but 6 cases. One EIKEN descriptor was removed from the set for further analysis; the other 62 were considered to be sufficiently aligned to the GSE.

Three types of analysis were carried out:

1. Rasch rating scale analysis

- To place all EIKEN and GSE descriptors on a single scale, all of them were concurrently calibrated
- To place the current data on the existing scale, both stability check (Taherbhai & Seo, 2013) and fixed equating method (Taherbhai & Seo, 2007) were applied

2. Correlation Analysis:

- After placing all the descriptors on a single scale, correlation analysis was conducted to quantify the degree to which EIKEN and GSE were related.

3. Regression Analysis

- To quantify the degree to which EIKEN levels (independent variable) would predict GSE values (dependent variable), regression analysis was conducted.

In more detail, step by step:

1. The following data cleaning criteria were applied to the data files (see 'flags' below):
   - Remove any descriptors with <0.8 certainty ratings.
   - Remove any raters with fewer than 25 ratings and point- biserial (individual rater vs. average of all raters) <0.10.
2. All descriptors (anchors, EIKEN and new GSE) were freely and simultaneously estimated to place all of them on the same scale.
3. The first round of free equating provided item infit and outfit indices:
   - Remove any items with infit mean square and/or outfit mean square >2.56.
   - This procedure indicated 4 GSE items and 1 EIKEN item; these were dropped from the next free calibration.
4. Item difficulties of the 66 anchor items were obtained from previous IRT runs (IRT_2 and IRT_3).

5 A stability check (Taherbhai & Seo, 2013) was conducted on the 66 anchor items. The stability result indicated that for all 66 items difficulty parameters should be regarded as fixed.

6 All descriptors were then re-run, fixing only the anchor items with the item difficulties obtained in Step 5 above.

7 The fixed equating method (Taherbhai & Seo, 2007) of Step 5 placed all the descriptors on the same scale as the previous analyses (IRT_2 and IRT_3).

8 The regression formula used in previous analyses was then applied to all the descriptors to place them onto the North 2000 scale:

Transformed_Item_Difficulty = 1.0508 * IRT_3 measure - 0.0308

9 The standard scaling formula as used in previous analyses was applied to the transform item difficulty of each descriptor in order to obtain the final GSE values: GSE = Transformed_Item_Difficulty * 7.804 + 54.

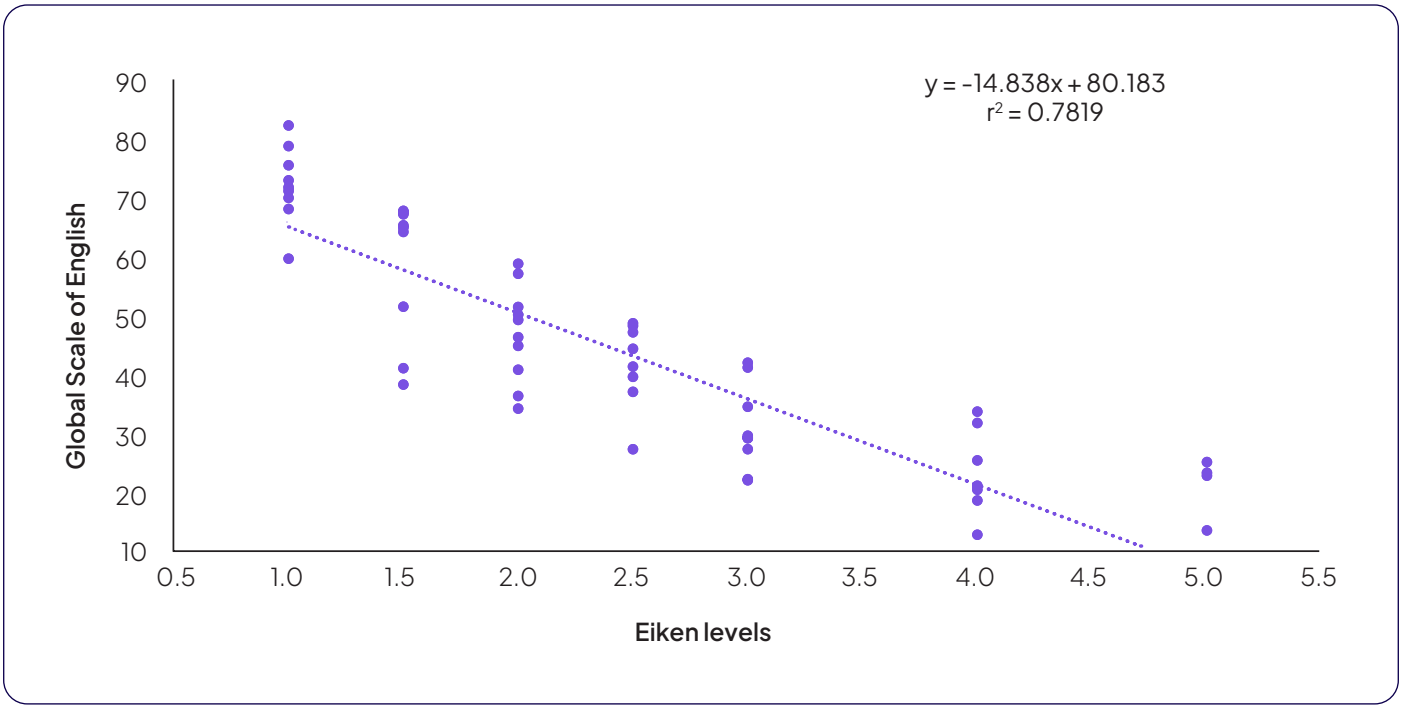
The full set of IRT results is to be found in Appendix 2. The following is a summary of the key findings:

- Certainty (inter-rater agreement) varied from 0.94 down to 0.55, but 57 descriptors had acceptable certainty (> 0.7).
- Mean-square fits (infit/outfit) range from 0.59 to 2.03, with 54 descriptors in the acceptable range 0.5 ..1.5 (and only 2 at > 2.0).
- Out-of-acceptable range values for these and other measurements (infit, outfit, and number of ratings) were flagged, and the number of flags counted. These flag counts were used to identify poorly performing descriptors, as follows: 0/1 Flag: no action, 2/3 Flags: check descriptor (identify possible rewrite or remove in the case of new Pearson descriptors), 4 Flags: rewrite, 5/6 Flags: remove.
- 41 EIKEN descriptors had 0 flags, 1 descriptor had 1 flag, and the remaining 21 had 2 flags. No EIKEN descriptors had more than two flags, and only one was removed (because of out-of range infit and outfit).

Correlation analysis was conducted to measure the degree to which the EIKEN and GSE were related. The result indicated that the GSE-EIKEN correlation coefficient was -0.90 ($p < .01$) (Note that the EIKEN scale, in contrast to the GSE, counts down as estimated proficiency level increases). It should also be noted that one of the EIKEN items ("Can recognize dates and days of the week.") performed as an outlier, and was dropped from further statistical analysis.

Regression analysis was conducted to quantify the degree to which EIKEN levels (independent variable) would predict the GSE values. As seen Figure 1, the r-squared of the equation is 0.78 ($p < 0.01$), and intercept and slope of independent variable are -14.84 and 80.18, respectively.

EIKEN Predicting GSE



Conclusions

EIKEN descriptors have been shown to be aligned to the GSE and (hence) to the CEFR. This testifies to the rigour with which the EIKEN Foundation have approached the specification and design of their tests, and to the accuracy of the standard-setting exercises carried out by Dunlea and Matsudaira (2009). This and other studies carried out by EIKEN only concerned the top EIKEN levels (which the present study did not examine). However, the present paper offers further implicit confirmation of the alignments obtained in these previous studies. The average certainty of EIKEN descriptors, at 0.788, was very close to that of the GSE Learning Objectives (CEFR originals and GSE new) rated in this and previous studies. This is particularly impressive, considering that the levels assigned to EIKEN descriptors are derived from learner self-assessment, whereas the levels assigned to GSE Learning Objectives are based on assessment by curriculum experts and experienced teachers.

The fact that EIKEN descriptors were successfully and accurately rated alongside GSE Learning Objectives suggests that further studies of this kind could be carried out, and that it might be possible for EIKEN descriptors to be included in the sets of GSE Learning Objectives - and GSE Learning Objectives could be used in future EIKEN test specifications. These results also imply that aligned EIKEN tests can report on the GSE scale to offer more precision to students, teachers and score users in reporting EIKEN test results. The alignment to the GSE and the CEFR can also be used to obtain wider recognition of EIKEN test results outside of the Japanese context.

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.
certainty value	The proportion of ratings within two adjacent categories on a categorical scale.
correlation	A statistic showing the interdependence between two variables.
explained variance (r-squared)	A statistical measure of how close the data are to the fitted regression line - in other words, how accurately the variances in the data is predicted.
GSE Learning Objective	A description of what a student is likely to be able to perform successfully 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: Application of a mathematical model to test data, which 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.
Rasch model	A psychometric model for analysing categorical data. It is a function of test taker ability and item difficulty, both placed on the same scale. It is a special case of the family of item response theory (IRT) models.

regression function	A mathematical function expressing the relation between a dependent variable (y-axis) and an independent variable (x-axis).
standard deviation	(SD) A statistic showing the amount of variation in a data-set. An SD close to 0 means all data points are close to the mean.
transformation function	A mathematical function enabling the transformation of values from one scale to corresponding values on another scale.
z-score (z-diff)	A statistical measure of a score's relationship to the mean in a group of scores, expressed in standard deviations of these scores to the mean. A z-score of 1 indicates a score at one standard deviation above the mean.

References

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

De Jong, J., Mayor, M. & Hayes, C (2016). *Developing Global Scale of English Learning Objectives aligned to the Common European Framework*

Pearson Global Scale of English Research Series. Accessible at https://prodengcom.s3.amazonaws.com/GSE-WhitePaper-Developing-LOs.pdf

Dunlea, J., & Matsudaira, T. (2009). *Investigating the relationship between the EIKEN tests and the CEFR*. In N. Figueras & J. Noijons (Eds.), Linking to the CEFR levels: Research perspectives. Arnhem, CITO and EALTA

EIKEN (2017) *Investigating the relationship of the EIKEN tests with the CEFR. EIKEN*. Foundation of Japan, Tokyo. Accessible at http://www.eiken.or.jp/eiken/en/research/

Linacre, J. M. (2009). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com

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

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut

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

Appendix 1: adapted EIKEN descriptors in the GSE Learning Objectives

Descriptor	Skill	EIKEN level	GSE	CEFR
Can write cardinal numbers from 1 to 20 as words.	Writing	Grade 3	21	<A1
Can introduce themselves in a basic way, giving some information about where they live, their family etc.	Speaking	Grade 3	27	A1
Can understand basic factual statements.	Listening	Grade 3	28	A1
Can express their emotions in a basic way.	Speaking	Grade Pre-2	30	A2
Can understand simple public announcements, provided that they are repeated (e.g. at airports, railway stations).	Listening	Grade 3	30	A2
Can write simple plans and arrangements on a calendar or in a diary.	Writing	Grade Pre-2	31	A2
Can give a simple description of their school or workplace.	Speaking	Grade 2	32	A2
Can ask a range of basic questions about colour, size, price etc. when shopping.	Speaking	Grade 2	32	A2
Can understand short, simple narrative texts.	Reading	Grade 3	36	A2+
Can write texts describing favorite objects, possessions or household pets.	Writing	Grade Pre-2	36	A2+
Can write short texts about their likes and dislikes, with explanations.	Writing	Grade 3	38	A2+
Can write simple texts giving key information about their culture (e.g. food, national holidays, festivals).	Writing	Grade Pre-1	42	A2+
Can understand the detail of weather forecasts.	Listening	Grade 2	42	A2+
Can describe in some detail what they are looking for when shopping.	Speaking	Grade 2	43	B1
Can use a range of simple fillers and interjections in conversation (e.g. “I see.” ”Right.”).	Speaking	Grade 3	46	B1
Can understand factual texts on general topics (e.g. guidebooks, history books).	Reading	Grade 2	49	B1
Can describe symptoms to a doctor in some detail.	Speaking	Grade Pre-1	49	B1
Can understand details of descriptions and explanations given by guides at tourist spots, museums, etc.	Listening	Grade Pre-1	49	B1
Can use a range of polite and informal expressions appropriate to different social contexts.	Speaking	Grade 1	53	B1+
Can follow classes and training courses on a range of subjects, provided the content is simple.	Listening	Grade 2	56	B1+
Can take notes to record the main points of lectures on familiar topics.	Writing	Grade 1	57	B1+
Can take notes to record the main points raised during meetings on familiar topics.	Writing	Grade 1	58	B1+
Can describe the plots of books or films in some detail.	Speaking	Grade Pre-1	61	B2
Can express their opinions in discussions on contemporary social issues and current affairs.	Speaking	Grade 1	61	B2
Can ask detailed questions in discussions on contemporary social issues and current affairs.	Speaking	Grade 1	72	B2+

Appendix 2: full IRT results for the EIKEN anchors

Eiken Descriptor	MEASURE	INMSQ	OUTMS	EIKEN	Certainty	Flags	Predicted GSE	GSE	Difference	z-diff	CEFR Level
Can understand a wide variety of movies and TV dramas.	2.109	1.29	1.46	1	0.78	1	65	71	6	0.63	B2+
Can understand instructions and announcements when taking flights and using public transportation (e.g. instructions for changing trains, announcements explaining delays).	-1.875	1.65	1.65	Pre-1	0.75	2	58	38	-20	-2.29	A2+
Can recognize dates and days of the week.	-5.590	1.66	1.65	5	0.68	2	10	10	0	-0.04	<A1
Can understand most TV news and current affairs programmes.	1.967	1.18	1.18	1	0.72	0	65	70	5	0.52	B2+
Can recognise phrases and content words related to basic personal and family information.	-4.039	1.23	1.24	4	0.77	0	21	21	0	-0.04	<A1
Can understand lengthy talks and monologues on a wide range of topics and issues (e.g. speeches and lectures for general educational purposes).	2.639	1.25	1.25	1	0.76	0	65	75	10	1.08	B2+
Can understand lengthy talks and monologues about topics that he/she is interested in (e.g. speeches, lectures).	1.701	0.84	0.85	Pre-1	0.80	0	58	68	10	1.08	B2+
Can understand important information when listening to public announcements (e.g. announcements paging people, information at events).	-0.900	1.01	1.01	2	0.85	0	51	46	-5	-0.60	B1
Can understand simple announcements (e.g. meeting place, arrival and departure times for transportation).	-3.206	1.17	1.17	Pre-2	0.81	0	43	27	-16	-1.84	A1

Eiken Descriptor	MEASURE	INMSQ	OUTMS	EIKEN	Certainty	Flags	Predicted GSE	GSE	Difference	z-diff	CEFR Level
Can understand the content of simple talks and monologues about familiar topics related to everyday life, provided that the speaker speaks slowly and/or repeats sections (e.g. school, club activities, talking about the weekend).	-2.331	1.07	1.07	3	0.88	0	36	35	-1	-0.15	A2
Can understand descriptions of the location of people and things (e.g. "The book is on the TV.").	-3.435	1.11	1.11	4	0.85	0	21	26	5	0.52	A1
Can recognize numbers used in familiar ways for everyday purposes (e.g. telephone numbers, the time, a person's age).	-4.876	1.58	1.58	5	0.65	2	10	14	4	0.41	<A1
Can understand simple explanations given by salesclerks when shopping (e.g. information about product sizes, discounts, whether a product is out of stock).	-2.369	1.72	1.70	2	0.70	2	51	34	-17	-1.95	A2
Can understand a speaker on the telephone, provided the content is simple (e.g. agreeing when to meet, taking short messages).	-2.024	0.64	0.64	Pre-2	0.83	2	43	37	-6	-0.72	A2+
Can understand simple work-related questions asked on phone calls.	-1.538	0.66	0.66	Pre-1	0.85	2	58	41	-17	-1.95	A2+
Can understand detailed meaning in extended conversations on familiar topics, if delivered in clear standard speech.	0.406	0.70	0.70	2	0.80	0	51	57	6	0.63	B1+
Can follow the main points of short talks on familiar topics if delivered in clear standard speech.	-1.138	1.03	1.04	Pre-2	0.74	0	43	44	1	0.07	B1
Can understand simple directions for how to get somewhere on foot, if spoken slowly and clearly and using a map.	-2.927	0.95	0.94	3	0.83	0	36	30	-6	-0.72	A2
Can recognise simple informal greetings.	-5.465	2.03	2.05	5	0.55	2	10	10	0	-0.04	<A1

Eiken Descriptor	MEASURE	INMSQ	OUTMS	EIKEN	Certainty	Flags	Predicted GSE	GSE	Difference	z-diff	CEFR Level
Can understand practical texts describing how to do things (e.g. cookbooks, gardening books).	-1.078	1.25	1.24	2	0.80	0	51	45	-6	-0.72	B1
Can identify specific information in a linguistically complex factual text.	2.186	1.03	1.02	1	0.76	0	65	72	7	0.75	B2+
Can understand complex technical information such as operating instructions, specifications for familiar products and services.	1.402	1.16	1.17	Pre-1	0.75	0	58	65	7	0.75	B2
Can scan several short informational texts on the same theme to find relevant information.	-0.438	1.14	1.14	2	0.73	0	51	50	-1	-0.15	B1
Can infer information from the labels on basic diagrams (e.g. bar charts, timelines) accompanying simple informational texts.	-0.799	1.44	1.44	Pre-2	0.72	2	43	47	4	0.41	B1
Can understand the main points in simple descriptive texts on familiar topics.	-1.524	0.75	0.75	3	0.83	0	36	41	5	0.52	A2+
Can find specific, predictable information in everyday materials (e.g. menus, timetables).	-2.666	1.30	1.31	4	0.74	0	21	32	11	1.20	A2
Can recognise the letters of the alphabet in upper and lower case.	-7.423	1.22	1.20	5	0.89	0	10	10	0	-0.04	<A1
Can understand works of literature (e.g. novels).	2.328	1.22	1.22	1	0.76	0	65	73	8	0.86	B2+
Can understand the main points of lengthy texts (e.g. required readings and materials for lectures and training courses).	1.267	0.61	0.61	Pre-1	0.85	2	58	64	6	0.63	B2
Can understand the main points of lengthy texts, provided the content is simple (e.g. required readings and materials for lectures and training courses).	0.611	0.95	0.96	2	0.86	0	51	59	8	0.86	B2
Can find streets, shops, and hospitals, etc., on simple maps written in English.	-3.822	1.54	1.55	3	0.60	2	36	22	-14	-1.62	A1

Eiken Descriptor	MEASURE	INMSQ	OUTMS	EIKEN	Certainty	Flags	Predicted GSE	GSE	Difference	z-diff	CEFR Level
Can understand simple signs and notices in public facilities (e.g. “No Smoking,” / “Closed” / “No Dogs”).	-5.734	1.66	1.65	4	0.67	2	21	10	-11	-1.28	<A1
Can understand basic sentences describing familiar activities from everyday life (e.g. “I play tennis every day.”).	-3.749	0.66	0.66	5	0.89	2	10	23	13	1.42	A1
Can participate in extended, detailed professional discussions and meetings with confidence.	3.033	0.78	0.77	1	0.83	0	65	79	14	1.53	C1
Can make a detailed, formal, evidence-based complaint about the quality of a product or service.	1.670	0.95	0.95	Pre-1	0.80	0	58	67	9	0.97	B2+
Can ask for, follow and give detailed directions.	-0.278	1.04	1.04	2	0.73	0	51	51	0	-0.04	B1+
Can make simple future arrangements and plans with reference to a diary or schedule.	-1.505	0.62	0.63	Pre-2	0.86	2	43	41	-2	-0.27	A2+
Can make and respond to suggestions.	-1.425	0.59	0.59	3	0.84	2	36	42	6	0.63	A2+
Can establish basic social contacts with simple, polite greetings and farewells.	-3.470	1.71	1.70	5	0.64	2	10	25	15	1.65	A1
Can negotiate over the telephone concerning a wide range of topics (e.g. making changes to schedules, negotiating prices).	1.738	1.17	1.17	1	0.73	0	65	68	3	0.30	B2+
Can ask questions and express opinions about the content of lectures and presentations, etc., concerning his/ her work or field of expertise.	1.630	0.81	0.82	Pre-1	0.78	0	58	67	9	0.97	B2+
Can explain familiar situations that occur in everyday life (e.g. explaining reasons for being late or absent).	-1.570	0.68	0.68	2	0.86	2	51	41	-10	-1.17	A2+
Can ask simple questions (e.g. the time, someone’s likes, name).	-3.956	1.11	1.10	4	0.70	0	21	21	0	-0.04	<A1

Eiken Descriptor	MEASURE	INMSQ	OUTMS	EIKEN	Certainty	Flags	Predicted GSE	GSE	Difference	z-diff	CEFR Level
Can answer yes/ no questions about familiar topics from everyday life (e.g. about likes and dislikes).	-3.691	1.09	1.08	5	0.81	0	10	23	13	1.42	A1
Can give dates using standard formats (day and month).	-4.260	1.04	1.03	4	0.74	0	21	19	-2	-0.27	<A1
Can talk about dreams and ambitions for his/her future (e.g. countries that he/she would like to visit, the career he/she would like to pursue).	-0.666	0.98	0.97	Pre-2	0.82	0	43	48	5	0.52	B1
Can describe routine actions from everyday life (e.g. "I got up at seven." / "I ate some bread for breakfast.").	-2.979	0.70	0.70	3	0.94	2	36	29	-7	-0.83	A1
Can write short, simple notices giving information about forthcoming events or activities (e.g. place, time, day) using appropriate layout, given a model.	-1.715	0.81	0.81	Pre-2	0.80	0	43	40	-3	-0.38	A2+
Can write a formal letter of complaint about products or services (e.g. about damaged products or unsatisfactory service).	0.716	0.89	0.89	1	0.83	0	65	60	-5	-0.60	B2
Can write short messages (e.g. "Ken called at 3 p.m.").	-2.971	0.82	0.82	3	0.86	0	36	29	-7	-0.83	A1
Can write a short diary entry (from one to three sentences).	-3.205	1.22	1.21	3	0.86	0	36	27	-9	-1.05	A1
Can write a research report including detailed analysis and evaluation of own and others' work on the topic of investigation.	3.458	1.03	1.05	1	0.79	0	65	82	17	1.87	C1
Can use fact and opinion effectively in writing.	1.355	0.82	0.82	Pre-1	0.80	0	58	65	7	0.75	B2
Can write simple personal recommendations related to cultural topics (e.g. books, films, concerts).	-0.544	0.59	0.59	2	0.83	2	51	49	-2	-0.27	B1
Can write a few basic sentences introducing themselves and giving basic personal information, given prompts or a model.	-3.844	0.85	0.85	3	0.83	0	36	22	-14	-1.62	A1

Eiken Descriptor	MEASURE	INMSQ	OUTMS	EIKEN	Certainty	Flags	Predicted GSE	GSE	Difference	z-diff	CEFR Level
Can use very basic connectors like and, but, so and then.	-2.432	0.77	0.76	4	0.87	0	21	34	13	1.42	A2
Can write their name, address and nationality.	-5.801	1.03	1.04	5	0.72	0	10	10	0	-0.04	<A1
Can describe what he/she would like to do and explain why (e.g. explain why he/she would like to study abroad or be employed by a certain company).	-0.272	0.78	0.78	Pre-1	0.82	0	58	52	-6	-0.72	B1+
Can write a simple description introducing the area in which he/she lives.	-2.110	0.75	0.74	2	0.84	0	51	36	-15	-1.73	A2+
Can write about his/her dreams and ambitions for his/her future (e.g. countries he/she would like to visit, the career he/she would like to pursue).	-0.614	0.65	0.65	Pre-2	0.86	2	43	49	6	0.63	B1
Can write dates and days of the week.	-4.965	1.14	1.13	4	0.76	0	21	13	-8	-0.94	<A1
Can print the letters of the alphabet (ABC...XYZ).	-6.873	2.03	2.10	5	0.86	2	10	10	0	-0.04	<A1


Be yourself
in English.