

Examining the difficulty pathways of can-do statements from a localized version of the CEFR

Judith Runnels

(Hiroshima Bunkyo Women's University, Japan)

jrunnels@h-bunkyo.ac.jp

(Received: 07.01.2013, Accepted: 02.02.2013)

Abstract

The Japanese adaptation of the Common European Framework of Reference (CEFR-J) is a tailored version of the Common European Framework of Reference (CEFR), designed to better meet the needs of Japanese learners of English. The CEFR-J, like the CEFR, uses illustrative descriptors known as can-do statements, that describe achievement goals for five skills (listening, reading, spoken production, spoken interaction and writing) across twelve levels instead of the CEFR's original six. The goal of the present analysis is to provide validity evidence in support of the inherent difficulty hierarchy within the 5 A level sub-categories (A1.1, A1.2, A1.3, A2.1 and A2.2) in two ways: 1) by testing whether the difficulty of the can-do statements for each skill increases with the levels, and 2) by determining if there are significant differences in difficulty ratings between each level. It was found that for most skills, the rank ordering from difficulty ratings made by Japanese university students somewhat matched the level hierarchy of the CEFR-J but that significant differences between many adjacent levels were not found. The localization of a general framework for use by a specific population of users and the limitations related to using a system of can-dos that is derived from estimates of difficulty are discussed.

Keywords: Common European Framework of Reference; CEFR-J; can-do statements; language skills; difficulty hierarchy.

Introduction

In Japan, there is presently a lack of consistency across the systems employed by Japanese primary, secondary and tertiary educational institutions for the measurement of proficiency and progress of English language learners. Negishi (2011) suggests that introducing a common language framework in Japan would allow for standardization in the field of foreign language learning and teaching. O'Dwyer and Nagai (2011) recommend the Common European Framework of Reference (CEFR, Council of Europe, 2001) given the previous success of its usage in Europe (North, Ortega, & Sheehan, 2010) and growing interest in the system outside of Europe

(Figueras, 2012). One of the goals of such a system is to provide learners and educators with a set of learner-centered performance scales that allow for standardized assessment of level (North, 2007). The CEFR measures learner proficiency and progress via illustrative descriptors that describe communicative competencies in five skills: listening, reading, spoken production, spoken interaction and writing (North, 2007). The descriptors progress from easy to more difficult over six levels of proficiency (Council of Europe, 2001) and each descriptor provides a self-sufficient criterion of achievement (Skehan, 1984). While this progression of difficulty has been continually validated in a European context

for the CEFR, regarding the inherent difficulty hierarchy of localized versions of the system, comparatively little research exists. Given the increasing interest in applying the CEFR outside of Europe, the process of developing alternate versions “to suit local needs and yet still relate back to a common system” (Council of Europe, 2001, p. 32) requires further study.

Research on the implementation of the CEFR in Japan began in 2008 at the Tokyo University of Foreign Studies (Tono & Negishi, 2012; Negishi, Takada & Tono, 2011). Illustrative descriptors, known as can-do statements, from DIALANG (Council of Europe, 2001, pp. 231-234) were administered to 360 Japanese university students. The purpose was to test if the rank ordering of difficulty by Japanese students, target users of the system, matched what was predicted by the CEFR. The statements were indeed found to order consistently. A further study by Negishi (2011) showed that over 80% of English language learners in Japan fell within the A level of the CEFR (also known as the Basic User level): the CEFR’s can-do statements did not appear to provide specific enough criteria for distinguishing effectively between the population’s span of language learners and development of an alternate version thus began (Negishi, 2011).

The Japanese adaptation of the CEFR (known as the CEFR-Japan or CEFR-J), increased the number of levels from the CEFR’s original six to twelve (by breaking down the four A and B levels into nine). Furthermore, all of the can-do statements were contextualized for Japanese learners (Tono & Negishi, 2012) and tested to ensure that the rank ordering of difficulty matched the predictions of the system (Negishi, 2011). However, the development of a scale is only the first step in implementing a system (North & Schneider, 1998) and due to the new divisions and statements, further research, such as ensuring that target users of the system behave similarly to the

participants of the initial development studies, is required. In terms of ensuring the difficulty hierarchy of the CEFR-J, little beyond describing the development process has been published (see Tono & Negishi, 2012; Negishi, Takada & Tono, 2011; Negishi, 2011).

A preliminary study by Runnels (2013) measured the rank ordering of difficulty by almost 600 university students on the CEFR-J’s A1 and A2 sub-levels. While there was no disordering in the levels found (with A1.1 being ranked the easiest and A2.2 being ranked the most difficult), the mean difficulty ratings frequently exhibited no significant differences from adjacent sub-levels. It was suggested that perhaps this was due to the sub-divisions being too great in number: splitting the A1 level into three sub-levels and the A2 level into two may limit the ability of users or assessors to be able to reliably distinguish features of language learners at each of those sub-levels (Runnels, 2013). On this, the Council of Europe (2001, p. 21) notes that “the number of levels adopted should be adequate to show progression...but should not exceed the number of levels between which people are capable of making reasonably consistent distinctions”. However, the lack of significant differences between levels in Runnels’ (2013) study may have been related to how the difficulties of each skill were being rated by participants in that perhaps one skill skewed the results of the entire level. Thus, the progression of difficulty should also be examined for each of the skills.

The current study was therefore designed to explore the difficulty pathways formed by difficulty ratings on can-do statements within each skill. Specifically, the inherent hierarchy of the CEFR (and the CEFR-J) requires that there be a gradual progression of easy to more difficult as a learner progresses up through the levels, and if this requirement is not met, the system’s intended function is lost. It is subsequently

expected that, like the levels, the skills should also order as predicted by the CEFR-J, with the A1.1 writing can-do statement, for example, being rated as more difficult than A1.2 writing and so on. It is not hypothesized that every skill will order perfectly, but a general tendency of increasing difficulty ratings across the levels for each skill is certainly expected. Furthermore, an ideal system might be one where the difficulty of A1.1 writing is comparable to A1.1 listening, with linear or exponential increases in difficulty between the levels, but the underpinning theory of the CEFR-J does not require this. What it does require, however, is that there are distinctions between the skills at each level (Council of Europe, 2001) and therefore, it is also hypothesized that significant differences in difficulty ratings between each level should exist. Ensuring this kind of a pathway means that the system is functioning as intended, and that the process of local contextualization of the system was successful.

Methods

Participants

590 first and second year students from a private university in Japan participated in this study. The survey was administered following completion of either one or three semesters of twice weekly 90 minute English classes. Participation was voluntary.

Instrument

The survey was administered on www.surveymonkey.com© (SurveyMonkey, 2012). Participants used a 5 point scale to indicate their perceived difficulty of the 50 randomly ordered, Japanese can-do statements from levels A1.1 to A2.2.

Procedure

For each CEFR-J level, there are 10 can-do statements (two for each of the five skills). The mean difficulty for each skill at each level (in logits) was calculated using Rasch measurement software Winsteps® (Linacre, 2010; for a full explanation of Rasch

analysis see Bond & Fox, 2007; Baghaei & Amrahi, 2011). To measure difficulty across levels within each skill, a logit difference of 0.3 is required for a significant main effect for difficulty (Miller, Rotou, & Twing, 2004; Lange, Greyson, & Houran, 2004).

Results

The following five figures illustrate the Rasch bubble pathways for each of the skills (Bond & Fox, 2007). Each level within the skill is represented with a circle, whose size is proportional to the standard deviation of the measure. The infit mean squares are shown on the x-axis where it can be seen that no items exhibit any misfitting infit (see Wright & Linacre, 1994). A larger value on the y-axis is associated with increased difficulty ratings.

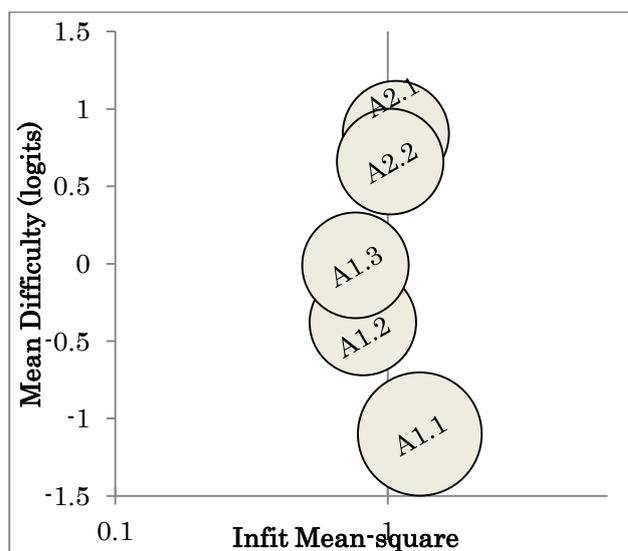


Figure 1: The bubble pathway for the mean difficulty of listening can-do statements across the CEFR-J's levels A1.1 to A2.2.

From Figure 1, it is evident that the ordering for the listening can-do statements for the A1 sub-levels was consistent with predictions of the CEFR-J, but that A2.2 falls below A2.1. The overall range of logits for all levels is 1.76. In terms of the logit difference required for a main effect of difficulty, the logit difference exceeds the required 0.3 difference for all adjacent

categories except for between A2.1 and A2.2.

The difficulty pathway for the reading can-do statements is shown in Figure 2. Some disordering is evident: the sub-levels from both A1 and A2 rated in the reverse direction of difficulty from what is predicted by the CEFR-J. Specifically, A1.3 is rated as less difficult than A1.2, and A2.2 as less difficult than A2.1. The span of logits is 0.91 and the required logit difference of 0.3 for significance exists between none of the adjacent categories except for between A1.3 and A2.1 although on this scale, these two levels do not fall adjacent to each other.

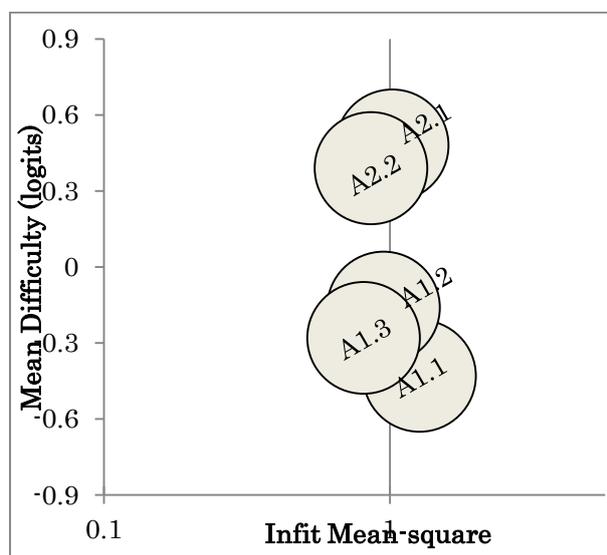


Figure 2: The bubble pathway for the mean difficulty of reading can-do statements across the CEFR-J's levels A1.1 to A2.2.

The spoken interaction pathway of difficulty ordered exactly as predicted by the CEFR-J (Figure 3). However, it is evident that the A1 sub-levels and A2.1 all fall very close to one another. Indeed, the range between all five levels spans only 1.04 logits. The only categories with a difference of greater than 0.3 logits are between categories A1.1 and A1.2 as well as A2.1 and A2.2.

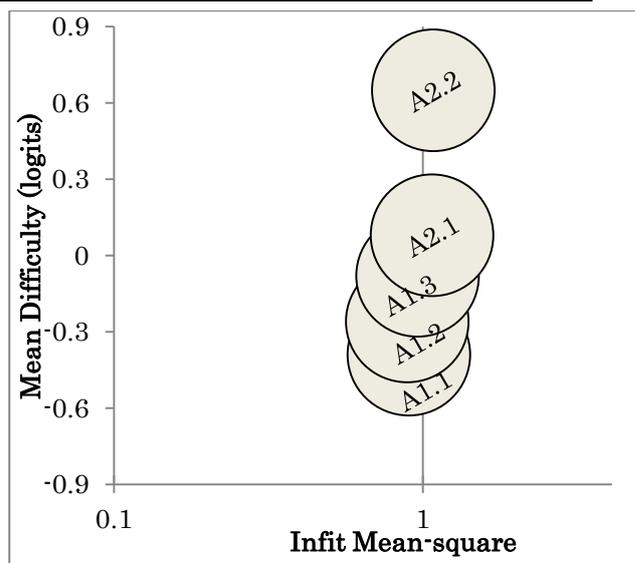


Figure 3: The bubble pathway for the mean difficulty of spoken interaction can-do statements across the CEFR-J's levels A1.1 to A2.2.

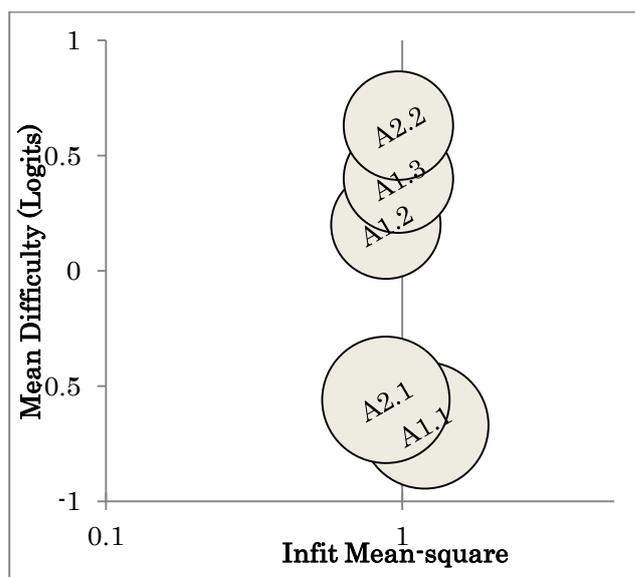


Figure 4: The bubble pathway for the mean difficulty of spoken production can-do statements across the CEFR-J's levels A1.1 to A2.2.

Figure 4 illustrates some major disordering of categories along the spoken production pathway of difficulty. Specifically, A2.1 has fallen below the difficulty ratings for A1.2 and A1.3 while A2.2 was rated as the most difficult. The span across all logit scores reaches 1.3.

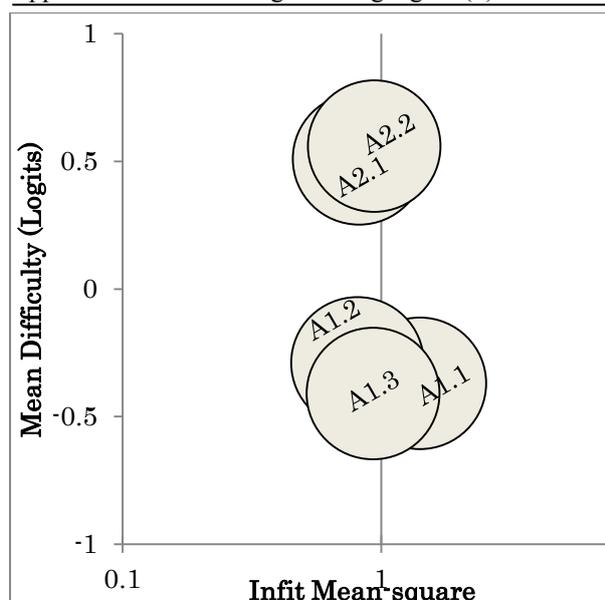


Figure 5: The bubble pathway for the mean difficulty of writing can-do statements across the CEFR-J's levels A1.1 to A2.2.

For the writing pathway shown in Figure 5, the can-do statements from both the A1 and A2 sub-levels grouped very closely together. The range of difficulty is only 0.97 logits and the 0.3 logit difference required for significance only exists between A1.3 and A2.1, or in other words, between the two higher order levels but not for any adjacent sub-levels.

To summarize the results of the rank ordering, the listening can-do statements performed reasonably well, with only the A2 sub-levels exhibiting disorder. Both the reading and spoken production can-do statements showed disordering at both the A1 and A2 levels whereas spoken interaction can-do statements ordered exactly as predicted. For writing, only the A2 sub-levels rank ordered as expected although the difference in difficulty ratings between the sub-levels at both the A1 and A2 levels is negligible.

In terms of the significant differences found between the levels for each skill, the listening can-do statements exhibited significant differences between all adjacent

A1 categories, but not for A2. For reading, the required significance level was found between only A1.3 and A2.1 (although due to disordering, these categories were not adjacent). Spoken production can-do statements behaved similarly, with no significant differences between any adjacent categories. While the spoken interaction can-do statements ordered as expected in terms of the CEFR-J, only the A2 sub-levels exhibited significant differences. Finally, for writing, differences between the higher-order A1 and A2 levels were evident, but not among the sub-levels.

Discussion

Overall, the difficulty judgments made by target users of the CEFR-J (Japanese university students) on can-do statements from A1.1 to A2.2 did not match entirely with the predictions of the CEFR-J. Moreover, most skills exhibited disordering and a lack of significant differences between adjacent categories was found for each skill. This relates to the preliminary findings by Runnels (2013) who found very little disordering overall, but a lack of significant differences between adjacent categories. It may be the case that performing this kind of an analysis on an individual skill's basis does not support the underpinnings of the CEFR-J which if language is seen as a uni-dimensional construct it should not be analysed modularly, according to skill. Nonetheless, the results herein suggest that the division of A1 and A2 into five sub-levels might be too great a number for users of the system to adequately and consistently distinguish features that are characteristic of learners at each level.

In fact, one of the major criticisms of the CEFR is that there is little empirical evidence to support the inherent hierarchy of increasing difficulty beyond the perception of language educators (Westhoff, 2007; Fulcher, 2003; 2004; 2010; Hulstijn, 2007) and it seems as if the participants in the current study perhaps do not share the same views as those of language educators.

In some cases, the contrasts between can-do statements across levels are quite subtle, as can be seen in the spoken interaction A1.2 (1) and A1.3 (2) statements where the primary difference is that the higher level A1.3 statement does not contain “using a limited repertoire of expressions”:

- (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.”
- (2) “I can ask and answer simple questions about familiar topics such as hobbies, club activities, provided people speak clearly.”

It may simply be that students do not associate an increase in difficulty between the requirements to complete such tasks in the same way that a language educator might. In fact, this highlights one of the major limitations of the current study and perhaps even of how the system was developed: the difficulty data is not comprised of scores on task performance. Rather, the analysis is based on difficulty judgments or self-assessment by learners. While the can-do statements are indeed designed to function as progress or proficiency markers when used by individual learners, the learners that did not associate less difficulty with the term “using a limited repertoire of expressions” may not behave the same way on a self-assessment, as they might during a more formal kind of performance-based assessment.

Nevertheless, the results also suggest that replications of the current study with other samples of student populations and at other CEFR-J levels might be useful in order to determine whether refinement or modification of the CEFR-J’s can-do statements and their level divisions is required. Alternatively, further contextualization of the existing can-do statements for use with the specific population of students, to ensure increasing

difficulty through the levels might also be necessary.

In either case, the CEFR-J is neither designed nor guaranteed to behave perfectly among every group of students or learners that is ever administered its can-do statements. In the current study, the hierarchy of difficulty was not consistently found which has implications for CEFR-J’s users: the scale of increasing difficulty is not always empirically supported (Westhoff, 2007; Fulcher, 2003; Hulstijn, 2007) and progression may proceed at differing rates or even in different directions for individual learners.

Conclusion

Ultimately, the results described herein highlight that the process of contextualization of a generalized European framework for local purposes outside of Europe is feasible and that the initial version of the CEFR-J’s levels and their illustrative descriptors was relatively successful. Indeed, developing and testing the CEFR is an on-going process involving both quantitative and qualitative methods, supplemented by replication studies (North, 2002; North, 2000; North & Schneider, 1998). Updates and modifications are continually being made. Although these processes are underway for the CEFR-J, additional empirical support is still required so that the CEFR-J can be used in the construction of curricula, materials and assessments for improving foreign language learning in the tertiary institutions of Japan or as a model for any organization looking to localize a general framework of reference.

References

- Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Baghaei, P., & Amrahi, N. (2011). Validation of a Multiple Choice

- English Vocabulary Test with the Rasch Model. *Journal of Language Teaching and Research*, 2, 1052-1060.
- Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences*. Mahwah NJ: Lawrence Erlbaum Associates.
- Figueras, N. (2012). The impact of the CEFR. *English Language Teachers Journal*, 66(4), 477-485.
- Fulcher G. (2003). *Testing second language speaking*. London: Longman/Pearson.
- Fulcher G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, 1(4), 253-266.
- Fulcher, G. (2010). The reification of the Common European Framework of Reference (CEFR) and effect-driven testing. *Advances in Research on Language Acquisition and Teaching: Selected Papers*, 15-26.
- Hulstijn J. A. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91(4), 663-667.
- Lange, R., Greyson, B., & Houran, J. (2004). A Rasch scaling validation of a 'core' near-death experience. *British Journal of Psychology*, 95, 161-177.
- Miller, G. E., Rotou, O., & Twing, J. S. (2004). Evaluation of the .3 logits screening criterion in common item equating. *Journal of Applied Measurement*, 5(2), 172-177.
- Negishi, M. (2011). CEFR-J Kaihatsu no Keii [The Development Process of the CEFR-J]. *ARCLE Review*, 5(3), 37-52.
- Negishi, M., Takada, T. & Tono, Y. (2011). A progress report on the development of the CEFR-J. *Association of Language Testers in Europe Conference*. Retrieved August 1st from: <http://www.alte.org/2011/presentations/pdf/negishi.pdf>.
- North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.
- North, B. (2002). Developing descriptor scales of language proficiency for the CEF common reference levels. In J.C.A. Alderson (Ed.), *Common European Framework of Reference for Languages: learning, teaching, assessment. Case studies*. Strasbourg: Council of Europe, 87-105.
- North, B. (2007). The CEFR Common Reference Levels: Validated reference points and local strategies. *Language Policy Forum Report*, 19-29.
- North, B., Ortega, A., & Sheehan, S. (2010). A core inventory for general English, *British Council/EAQUALS*. Retrieved August 3rd from: <http://www.teachingenglish.org.uk/publications/british-council-eaquals-core-inventory-general-english>.
- North, B. & Schneider, G. (1998): Scaling descriptors for language proficiency scales. *Language Testing*, 15(2), 217-262.
- O'Dwyer, F., & Nagai, N. (2011). The actual and potential impacts of the CEFR on language education in Japan. *Synergies Europe*, 6, 141-152.
- Runnels, J. (2013). Preliminary validation of the A1 and A2 sub-levels of the CEFR-J. *Shiken Research Bulletin*, in press.
- Skehan, P. (1984). Issues in the testing of English for specific purposes. *Language Testing*, 1(2), 202-220.
- SurveyMonkey. (2012). *SurveyMonkey.com*, LLC. Palo Alto, California, USA: www.surveymonkey.com.

- Tono, Y., & Negishi, M. (2012). The CEFR-J: Adapting the CEFR for English Language Teaching in Japan. *Framework & Language Portfolio SIG Newsletter*, 8, 5-12.
- Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *The Modern Language Journal*, 91(4), 676 – 679.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.