

**This is an electronic reprint of the original article.  
This reprint *may differ* from the original in pagination and typographic detail.**

**Author(s):** Alanen, Riikka; Huhta, Ari; Martin, Maisa; Tarnanen, Mirja; Mäntylä, Katja; Kalaja, Paula; Palviainen, Åsa

**Title:** Designing and Assessing L2 Writing Tasks Across CEFR Proficiency Levels

**Year:** 2012

**Version:**

**Please cite the original version:**

Alanen, R., Huhta, A., Martin, M., Tarnanen, M., Mäntylä, K., Kalaja, P. & Palviainen, Å. (2012). Designing and Assessing L2 Writing Tasks Across CEFR Proficiency Levels. In: Torrance, M., Alamargot, D., Castelló, M., Ganier, F., Kruse, O., Mangen, A., Tolchinsky, L., & Van Waes, L. (Eds). Learning To Write Effectively: Current Trends In European Research; Chapter 2.03.02. (Studies in Writing, Volume 25), Brill, pp. 221-223.

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

CHAPTER 2.03.2. DESIGNING AND ASSESSING L2 WRITING TASKS  
ACROSS CEFR PROFICIENCY LEVELS

Riikka Alanen, [riikka.alanen@jyu.fi](mailto:riikka.alanen@jyu.fi)

Ari Huhta, [ari.huhta@jyu.fi](mailto:ari.huhta@jyu.fi)

Maisa Martin, [maisa.martin@jyu.fi](mailto:maisa.martin@jyu.fi)

Mirja Tarnanen, [mirja.tarnanen@jyu.fi](mailto:mirja.tarnanen@jyu.fi)

Katja Mäntylä, [katja.mantyla@jyu.fi](mailto:katja.mantyla@jyu.fi)

Paula Kalaja, [paula.kalaja@jyu.fi](mailto:paula.kalaja@jyu.fi)

Åsa Palviainen, [asa.palviainen@jyu.fi](mailto:asa.palviainen@jyu.fi)

University of Jyväskylä

The CEFLING project (<http://www.jyu.fi/cefling>), funded by the Academy of Finland (2007-2009), was set up in 2006 for the explicit purpose of combining the research perspectives of language testing and second language acquisition. In the project, written L2 data were collected from young (aged 12-16) and adult L2 Finnish and L2 English. The overall aim of the project is to describe the features of language that L2 learners use at various levels of language proficiency as described by the CEFR scales.

Data elicitation, or what kinds of tasks are used to collect L2 data is a critical issue for both SLA and language testing research. Languages are learned by using them in various tasks inside or outside the classroom both in formal and informal contexts. Tasks in this very broad sense provide language learners/users opportunities for language learning. Language task is defined here as “an activity which requires learners to use language, with emphasis on meaning, to attain an objective” (Bygate, Skehan & Swain, 2001).

L2 data used in CEFLING come from two sources. First, the adult L2 data come from the National Certificate language examinations, a national language testing

system run by the Centre for Applied Language Studies at the University of Jyväskylä. The tasks and the data from adult language learners have yet to be analyzed systematically. Secondly, L2 data were also collected from 7<sup>th</sup> – 9<sup>th</sup> graders at school. A considerable amount of time and effort went into designing tasks suitable for eliciting L2 data from younger language learners. In what follows we will present an account of the procedures that went into designing and assessing these particular L2 writing tasks.

Our first requirement was that the tasks should be communicative and that they should have some measure of authenticity in terms of text types and processes needed in completing the tasks. A number of tasks used in the National Certificate language examination, which have been found reliable and valid for testing the proficiency of adult L2 learners, were reviewed to find a number of suitable candidates for data elicitation. Secondly, it was decided that task types themselves should be familiar to students so that they would find it easier to understand what was required of them. For this purpose, language course books used at school were reviewed. Finally, it was decided that the tasks should be aimed at the appropriate L2 proficiency level, that is, they should be neither too easy nor too difficult: L2 users/learners in this instance were estimated to be at least at the level A1 and at B2 at the most.

Based on such considerations, a set of communicative writing tasks were designed for the project. The tasks were piloted by asking a small number of students to complete them. The results were then analyzed and some changes were made as to the number of tasks and instructions. The final set of tasks consisted of five communicative tasks representing a variety of functions, text types and register.: Task 1 was an informal email message to a friend, Task 2 was an informal email message to a teacher, Task 3 was a formal message to an internet store, Task 4 was an opinion piece, and Task 5 was a story. Additional tasks were designed to elicit data on particular linguistic features (such as questions, relative clauses, and word derivation) in L2 English.

Assessing the L2 learners' performance on the tasks was carefully planned, as well. The assessment criteria and procedures necessary for the reliable and valid rating of the tasks and task performances based on CEFR were prepared. The criteria used in assessment consisted of relevant CEFR scales for writing, including scales for written interaction, correspondence and notes, messages and forms, creative writing and thematic development, and overall written production. A team of raters for L2 Finnish and L2 English was selected and trained for the project. The raters were also given a set of benchmarks selected from the pilot performances; in the case of L2 English, they were also given benchmarks prepared by the Council of Europe.

A total number of 3427 performances were collected from 7<sup>th</sup> to 9<sup>th</sup> graders from schools. A number of scripts (N = 1789) were selected for assessment by a team of trained raters (N=9 for English, N = 4 for Finnish). Each script was rated by four (L2 English) or three (L2 Finnish) trained raters. Only those scripts were selected for the analysis of linguistic features on which the raters were in sufficient agreement (three out of four raters for L2 English, two out of three for L2 Finnish). Thus, the final corpus for L2 English contained 562 scripts and for L2 Finnish 825 scripts.

The suitability of the tasks for users/learners at various CEFR levels was assessed by raters before data collection began. These assessments were then compared with the ratings of the L2 learner performances. The findings so far indicate that the raters were for the large part in agreement about the difficulty of the tasks: for example, A2 was deemed within the range of all of the tasks. Similarly, various statistical analyses (e.g. multifaceted Rasch analyses) of the L2 users/learners' performance across the five communicative tasks reveal little variation in task difficulty across the five tasks (Alanen & Huhta 2009). Overall, these findings indicate that the task design was quite successful in eliciting L2 data across the proficiency levels from A1 to B2, but rather less so for advanced levels; however, this was only to be expected because of the young age of the participants. In sum, CEFLING is one of the few projects to date in which an attempt has been made to combine the methodological rigor of a theory-bound SLA research with an understanding of the need to define L2 performance according to a strictly defined set of standards based on a model of communicative language use.

- Alanen, R. & Huhta, A. (2009). L2 learners' performance across L2 writing tasks: comparing tasks and language proficiency through CEFR scales. Paper at TBLT 2009, 13-16 September 2009, Lancaster, UK.
- Bygate, M., Skehan, P. & Swain, M. (eds.) 2001. *Researching pedagogic tasks. Second language learning, teaching and testing.* Harlow: Longman / Pearson Education.