Research Note: Diagnosing the development of writing ability

Ari Hutha
University of Jyväskylä, Finland

March 2014

1. Introduction

The study reported on here was integrated into the larger DIALUKI study (Diagnosing Reading and Writing in a Second or Foreign Language; see www.jyu.fi/dialuki), the first part of which was a relatively large-scale cross-sectional exploratory study of how reading and writing in the first and foreign/second language (L2) relate to each other and how a range of other language skills, motivational factors, background variables and cognitive/psycholinguistic variables relate to reading and writing in L1 and L2 (see also Alderson, Haapakangas, Huhta, Nieminen & Ullakonoja, in print). The study was carried out in Finland in 2010-11.

The main aims of the study reported here were:

- To examine the relationship between the overall ability to write in the first language (Finnish) and a foreign language (English).
- To examine the relationship between the reading and writing ability in English as L2.
- To examine the relationship between writing/reading in L2 English and writing/reading in L1 Finnish.
- To identify cognitive and personal features contributing to the ability to perform well on tests of L2 English writing.
- To find out if the results related to the above points differed in any systematic way between the two age groups studied (14-year-olds vs 17-year-olds).

In addition, the study examined the relationship between the rating systems used in the DIALUKI study and in the rating of writing performances in the Pearson Test of English General (PTE General), the details of which will be published in a separate research article (Huhta & Lamprianou, forthcoming).

2. Research design

Informants: Two groups of Finnish learners of English as a foreign language were studied: 8th graders in the comprehensive school (14-year-olds) and second year gymnasium students (17-year-olds). The students were volunteers coming from about 10 different comprehensive schools and 10 gymnasia in different parts of

the country. They were not a statistically representative sample of learners of English in those grades but nevertheless represented a range of different regions (e.g., towns and rural communities), sizes of schools, and ability levels. By the time of the study the 8th graders had studied English for over six years (since grade 3) and the gymnasium students for over nine years (also from grade 3). The comprehensive schools include all kinds of students as they are part of compulsory education, whereas the gymnasia (upper secondary schools) are for those students who wish to pursue more academically oriented studies and perhaps enter tertiary education (typically, somewhat over 50% of the age group go to a gymnasium, rather than enter vocational secondary education, or go directly to work). A total of 202 8th graders and 195 gymnasium students completed at least one of the two PTE General writing tasks used in the study. The larger DIALUKI study also included a younger group of English learners and groups of learners of Finnish as L2 but these were not part of the study reported here.

Tests. The informants responded to a large number of different tests and questionnaires that tapped a range of L1 and L2 language skills, psycholinguistic skills, as well as motivation and background characteristics. Informants' parents also replied to a background questionnaire. The test battery included English reading and writing tasks from operational PTE General tests. The two groups of learners involved in this study completed two writing tasks and a total of 25 reading items. In addition, they took one local English writing task (from the project CEFLING, see Alanen, Huhta, Jarvis, Martin & Tarnanen, 2013, and www.jyu.fi/cefling) and a full computerized DIALANG English reading test. The two PTE writing tasks taken by the 8th graders were at the A2 (a Section 8 task in PTE General) and at the B1 (Section 9) levels of the CEFR. The gymnasium students took a B1 (Section 8) and a B2 (Section 9) writing task. Some of the writing tasks were so-called independent tasks, i.e., they did not relate to any other tasks, whereas others were integrated tasks, i.e., they related to a reading task that the students had just taken. In an integrated writing task, the task was on the same theme as the reading task and the students were expected to relate, in some general way, what they wrote to the text they had just read. The PTE General reading tasks covered Sections 4, 5, 6 and 7 in both target groups. The 8th graders did tasks that were estimated to be at levels A2 and B1 and the gymnasium students did tasks at levels B1 and B2.

Rating of writing. The PTE General writing tasks were rated by two teams of raters, one in Finland and another in the UK. Each script was double rated both in Finland (by DIALUKI raters) and in the UK (by Pearson raters). The rating in Finland was carried out by using a rating scale that combined several writing scales taken from the CEFR. The scripts were, thus, rated directly on the CEFR levels. This procedure had successfully been used earlier in the CEFLING project and it was also used in DIALUKI (see Alanen, Huhta & Tarnanen, 2010, and Huhta, Alanen, Tarnanen, Martin & Hirvelä, forthcoming). The Pearson raters used Pearson's own CEFR-related criteria and procedures. The main difference between the two approaches is that the Finnish raters did not make any assumptions about the CEFR level that a particular writing task was pitched at, and therefore, a performance could be awarded anything between A1 (or even below A1) and C2. In contrast, in the Pearson rating system there is an assumption that each writing task elicits writing ability at a particular CEFR level, and therefore, the rating is about judging if the learner's performance is at the intended CEFR level or whether it falls below it or whether it exceeds the level. More precisely, the PTE General writing tasks were rated with a 5-point scale where 1 means that the performance is below the intended level, 2, 3 and 4 refer to performance at the

level (2 being close to the level below and 4 the level above), and 5 means that the performance exceeds the requirements of the intended level.

The following table summarises the rating design:

Writing task	Dataset / Level	Rating in Finland	Rating in the UK
PTE General A2 task (independent task)	8th graders	2 raters (out of a pool of 11 raters)	2 raters (out of a pool of 5 raters)
PTE General B1 task (integrated task)	8th graders	as above	as above
PTE General B1 task (independent task)	gymnasium	as above	as above
PTE General B2 task (integrated task)	gymnasium	as above	as above
CEFLING English task (independent task)	8th graders gymnasium	as above	N/A

Data analyses. In addition to computing descriptive statistics and carrying out classical item analyses, the L1 and L2 reading test scores were analysed with the WinstepsRasch analysis software and the ratings of the L1 and L2 writing performances with the Facets programme (Multifaceted Rasch analysis).

3. Findings

Students' CEFR levels in writing and reading in English. As could be expected, the gymnasium students performed better; their mean writing level was B1 on both tasks whereas the 8th graders were one CEFR level below them (at A2). This is in line with previous national assessments of English in Finland. Unfortunately, we did not have information available that would have enabled us to relate the scores based on the PTE General reading items used in the study with the CEFR levels. However, as one 5-item PTE General reading task (intended for B1) was the same for both groups, we could directly compare the two groups and see that the 8th graders' average score on that task was 3.6 (about 70% of the total score) whereas the gymnasium students scored 4.5 on average (about 90%).

Correlations between reading and writing in L1 and L2. Reading and writing in English correlated quite strongly with each other: in both student groups the correlations ranged from slightly below .5 to almost .7. Ability to read in L1 Finnish correlated with reading in English, but the correlations were fairly modest in both groups (about .3 to .4). Writing in L1 Finnish had modest .2 to .3 correlations with ability to write in English.

Characteristics of the ratings and rating scales. The Facets analyses indicated that while there were differences in the rater severity in both rater groups, none of the raters was misfitting. The analyses also indicated that both types of rating scales appeared to function as measurement instruments in the sense that the scale points were in the right order and separable from each other when used for

rating by the two groups of raters.

Comparisons of the ratings in Finland and in the UK. The comparison of the Finnish and Pearson ratings of the same performances demonstrate that there was a high but not a perfect correlation between the ratings of the two rater groups that used different approaches to CEFR-related rating. The correlations between the Rasch measures calculated from the Finnish and UK ratings were, for the 8th graders, .824 (Pearson correlation) / .817 (Spearman correlation), and for the gymnasium students, .783 / .776

For a different comparison of the outcomes of the two kinds of CEFR-based ratings, the Pearson 5-point ratings were converted to the CEFR scale in the following way. We take a B1 task as an example. Since the meaning of the scale points in the Pearson rating scale is related to CEFR levels in the way described earlier, the ratings of performances on a B1 task could be converted to CEFR levels by recoding 1 as A2, 2, 3 and 4 as B1 and 5 as B2. A similar conversion was carried out on the ratings of students' performances on the A2 and B2 tasks. This conversion is not exact but it gives us a fairly good approximation that enables us to compare the two ratings systems (and the actual ratings awarded by the raters) with reference to the CEFR levels. Obviously, there was no need to convert the ratings of the Finnish team as they gave their ratings directly on the CERF scale. To be precise, (rounded) fair averages from the Facets analyses rather than raw ratings were used in the above comparisons.

There was considerable overlap in the levels awarded by the two groups of raters for the 8th graders: about 85% of the ratings (i.e., CEFR levels awarded to the students) matched. The main difference between the ratings was that over 10% of the learners who had been given B2 by the Finns using the CEFR scale directly were left at B1 by the Pearson raters using their own, task-based approach to CEFR-related rating.

The correspondence of ratings was somewhat smaller in the gymnasium data but high nevertheless as about 70% of the ratings matched. This time, however, the direction of the difference was different from the 8th graders' ratings: it appeared that it was more difficult to be placed at higher CEFR levels if the learner was rated by the Finnish raters using the CEFR scale directly. In over 10% of the cases, learners who were rated at A2 by the Finns were placed at B1 or higher by the Pearson raters. Similarly, over 10% were rated at B1 by the Finns but at B2 by the Pearson raters in the UK (for more details, see Huhta & Lamprianou, forthcoming).

Predicting writing in English from psycholinguistic tasks. In the regression analyses, the psycholinguistic tasks used in the study accounted for 21% - 40% of the variance in students' performance on the English writing tasks. The best psycholinguistic predictor of writing in English across both types of ratings (the direct CEFR rating used by the Finnish raters and the task-based CEFR rating used by the Pearson raters) was the speed with which the learner could name (say aloud) in English a set of different colours, numbers and letters (a Rapidly Alternating Stimulus task). The other significant predictors included: the same naming task in L1 Finnish, a task requiring the recognition of rapidly presented words in L1 or L2, speed of reading a list of words in L1, and a working memory task in L2 (a backward digit span task based on numbers).

Predicting writing in English from motivation. Motivation turned out to predict L2 English writing performance somewhat better than the psycholinguistic variables. Across the two age groups and two kinds of writing indicators (Finnish

and Pearson ratings), motivation measures explained usually over 40% of the variance in writing in English. The best predictor was always the English self-concept which is about the learner's idea of him/herself either as a good or a poor learner or master of English. Also anxiety - or lack of anxiety, in fact - was another significant predictor (anxiety may not actually be part of motivation but a related construct).

Predicting writing in English from other linguistic skills. Performance on other linguistic measures was an even stronger predictor of writing in L2 English than motivation and psycholinguistic tasks. Linguistic measures explained from 50% to over 70% of variance in L2 writing performance. The consistently best predictor of writing in English was dictation in English. Several other linguistic variables turned out to be significant predictors, too, such as reading in L1, vocabulary in L2, segmentation (identifying word and sentence boundaries in a text) in L1 or L2, and writing in L1, as well as self-assessment of reading and writing in L2.

Predicting writing in English from learners' background characteristics. Ability to write in L2 English could be predicted from background to some extent: 20% - 32% of the variance could be explained by learners' background. The strongest predictor varied but typically it was the age at which the child had learned to read in L1 (according to the parents' memory) and how often the learner read in English outside school. Also the number of different languages that the learner reported knowing to any degree was among significant predictors, as was the parents' self-assessed command of English.

4. Conclusions

To conclude, we return to the main aims of the study, which relate to exploring the relationships between reading and writing in L1 Finnish and L2 English (the latter was measured by PTE General tests) and to predicting writing in L2 English with a range of linguistic, motivational, background and psycholinguistic variables. A comparison of the different CEFR-related rating systems used in Finland and in the PTE General was also of interest in the study.

It was clear that the ability to read and write in L1 Finnish was related to the learners' ability to read and write in L2 English in both age groups involved in the study but the relationship was not very strong. As has often been found in studies of L2 proficiency, reading and writing in L2 English were quite strongly associated with each other. The relationships between these L1 and L2 skills was quite similar in the 14-year-olds' and the 17-year-olds' groups. The main, and entirely expected, difference between the two groups was that the older group, which had studied English longer, performed better on the English reading and writing tasks (roughly A2 vs B1 on average).

The best predictors of performance on L2 English writing tasks were tests of other English skills: dictation (which involves writing), vocabulary, reading, and segmentation. From the diagnostic point of view, vocabulary and segmentation are more interesting than the more integrated skills of dictation and reading (see Alderson, 2005). Motivation also correlated with L2 writing but the strongest predictors (image of oneself as a learner of English and anxiety) are probably not causes of high or low writing performance. They are likely to have a reciprocal relationship with the L2 English abilities: good performance in English probably improves one's image of oneself as a (good) learner and user of English and decreases anxiety to use and study English, and vice versa. However, it is impossible to say anything about any causal relationships between these aspects

of motivation and language skills simply on the basis of correlational studies such as this one. Certain psycholinguistic and background features were also correlated with L2 writing ability although not often very strongly. However, some of the psycholinguistic tasks may have diagnostic value as they tap several underlying skills and processes involved in language use.

Finally, it was discovered that the two rather different rating systems used by the DIALUKI raters in Finland (a direct rating onto CEFR levels) and by the Pearson raters in the UK (a task-based CEFR-related rating) produced rather comparable results: a large majority of the performances were rated on the same CEFR levels. However, there were some systematic differences for 10%-20% of the ratings which suggest that the rating systems may not always produce exactly the same results. This remains to be examined in more detail in the future.

References

Alanen, R., Huhta, A.& Tarnanen, M. (2010). Designing and assessing L2 writing tasks across CEFR proficiency levels. In Bartning, I., Martin, M. & Vedder, I. (eds.) Communicative proficiency and linguistic development: intersections between SLA and language testing research. EUROSLA MonographSeries, 1. Pp. 21- 56.

Alanen, R., Huhta, A., Jarvis, S., Martin, M. & Tarnanen, M. (2012). Issues and challenges in combining SLA research and language testing. In Tsagari, D. &Csepes, I. (Eds.), Collaboration in language testing and assessment. Language Testing and Evaluation Series, Grotjahn, R. &Sigott, G. (general eds). Frankfurt am Main: Peter Lang.

Alderson, J.C. (2005). Diagnosing foreign language proficiency: the interface between learning and assessment. London: Continuum.

Alderson, J.C., Haapakangas, E-L., Huhta, A., Nieminen, L. & Ullakonoja, R. (forthcoming in 2014). Diagnosing Reading in a Second or Foreign Language. Routledge.

Huhta, A., Alanen, R., Tarnanen, M., Martin, M. & Hirvelä, T. (forthcoming in 2014). Assessing learners' writing skills in a SLA study - Validating the rating process across tasks, scales and languages. Language Testing 31, 3.

Huhta, Ari & Lamprianou, Iasonas. (forthcoming). Comparability of two CEFR-based rating systems.